

# On the von Bahr–Esseen inequality

Iosif Pinelis\*

*Department of Mathematical Sciences  
Michigan Technological University  
Houghton, Michigan 49931, USA  
E-mail: [ipinelis@mtu.edu](mailto:ipinelis@mtu.edu)*

**Abstract:** The well-known von Bahr–Esseen bound on the absolute  $p$ th moments of martingales with  $p \in (1, 2]$  is extended to a large class of moment functions, and now with a best possible constant factor (which depends on the moment function). As an application, measure concentration inequalities for separately Lipschitz functions on product spaces are obtained. Relations with  $p$ -uniformly smooth and  $q$ -uniformly convex normed spaces are discussed.

**AMS 2000 subject classifications:** Primary 60E15, 60B11, 62G10; secondary 46B09, 46B20, 46B10.

**Keywords and phrases:** probability inequalities, sums of independent random variables, martingales, v-martingales, concentration of measure, separately Lipschitz functions, product spaces,  $p$ -uniformly smooth normed spaces,  $q$ -uniformly convex normed spaces.

## Contents

1	Summary and discussion . . . . .	2
1.1	Summary . . . . .	2
1.2	Discussion . . . . .	3
1.2.1	Structure of the class $\mathcal{F}_{1,2}$ and examples of functions in this class . . . . .	4
1.2.2	On the best possible constant $C_f$ in general and, in particular, for the power and extreme functions . . . . .	6
1.2.3	Application: concentration inequalities for separately Lipschitz functions on product spaces . . . . .	9
1.2.4	Other corollaries of Theorem 1.1 and comparisons with known results . . . . .	11
2	Proofs . . . . .	15
2.1	Proofs of Propositions 1.2, 1.3, 1.4, 1.6, and 1.16 . . . . .	16
2.2	Proof of Proposition 1.8 . . . . .	20
2.3	Proofs of Corollary 1.9 and Propositions 1.10 and 1.11 . . . . .	24
2.4	Proof of Theorem 1.1 . . . . .	27
	References . . . . .	33

---

\*Supported by NSF grant DMS-0805946

## 1. Summary and discussion

### 1.1. Summary

In this subsection, we shall define a class of random variables (r.v.'s) and a class of generalized moment functions, for which a sharp probability inequality will be stated. This inequality is the main result of the paper. Originally, extension to a wider class of moment functions was not an objective of this study. Rather, the aim (suggested by statistical applications considered in [53]) was to obtain an optimal version of the von Bahr–Esseen (vBE) inequality [60] for the (absolute) power moments. However, it turned out that such an extension to general moment functions provided the only apparently available way to prove the best possible bound for the power moments.

Given any sequence  $(S_j)_{j=1}^n$  of (real-valued) r.v.'s, let  $X_j := S_j - S_{j-1}$  denote the corresponding differences, for  $j \in \overline{1, n}$ , with the convention  $S_0 := 0$ , so that  $X_1 = S_1$ ; here and in what follows, for any  $m$  and  $n$  in the set  $\{0, 1, \dots, \infty\}$  we let  $\overline{m, n}$  stand for the set of all integers  $i$  such that  $m \leq i \leq n$ .

If  $E|X_j| < \infty$  and  $E(X_j|S_{j-1}) = 0$  for all  $j \in \overline{2, n}$ , let us say that the sequence  $(S_j)_{j=1}^n$  is a *v-martingale* (where “v” stands for “virtual”); in such a case, let us also say that  $(X_j)_{j=1}^n$  is a *v-martingale difference sequence*, or simply that the  $X_j$ 's are v-martingale differences. Note that, for a general v-martingale difference sequence  $(X_j)_{j=1}^n$ ,  $X_1$  may be any r.v. whatsoever; in particular, its mean (if it exists) may or may not be 0. It is clear that any martingale  $(S_j)_{j=1}^n$  is a v-martingale. Quite similarly one can define v-martingales with values in a normed space.

Introduce the following class of generalized moment functions:

$$\begin{aligned} \mathcal{F}_{1,2} &:= \{f \in C^1(\mathbb{R}): f(0) = 0, f \text{ is even,} \\ &\quad f' \text{ is nondecreasing and concave on } [0, \infty)\} \\ &= \{f \in C^1(\mathbb{R}): f(0) = 0, f \text{ is even,} \\ &\quad f'' \text{ is nonnegative and nonincreasing on } (0, \infty)\}; \end{aligned} \quad (1.1)$$

here, as usual,  $C^1(\mathbb{R})$  is the class of all continuously differentiable real-valued functions on  $\mathbb{R}$ , and then  $f''$  denotes the right derivative on  $(0, \infty)$  of  $f'$ ; on  $(-\infty, 0)$ ,  $f''$  will denote the left derivative of  $f'$ . It is clear that each function  $f \in \mathcal{F}_{1,2}$  is convex and hence nonnegative. Also, for each function  $f \in \mathcal{F}_{1,2}$  one has  $f'(0) = 0$ . It follows that  $f' > 0$  on  $(0, \infty)$  and hence  $f > 0$  on  $\mathbb{R} \setminus \{0\}$  for any function  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ .

#### Theorem 1.1.

(I) For any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ ,  $n \in \overline{2, \infty}$ , and v-martingale  $(S_j)_{j=1}^n$ ,

$$E f(S_n) \leq E f(X_1) + C \sum_{j=2}^n E f(X_j) \quad (1.2)$$

with  $C = C_f$ , where

$$C_f := \sup_{0 < x < s < \infty} \frac{L_{f;s}(x)}{f(s)}, \quad (1.3)$$

$$L_{f;s}(x) := f(x - s) - f(x) + sf'(x). \quad (1.4)$$

(II) The constant factor  $C_f$  is the best possible in the sense that, for each  $f \in \mathcal{F}_{1,2} \setminus \{0\}$  and each  $n \in \overline{2, \infty}$ , the number  $C_f$  is the smallest value of  $C$  such that inequality (1.2) holds for all  $v$ -martingales  $(S_j)_{j=1}^n$ ; in fact,  $C_f$  is the best possible even if the differences  $X_1, \dots, X_n$  are assumed to be any independent zero-mean r.v.'s.

(III) For each  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ ,

$$1 \leq C_f \leq 2. \quad (1.5)$$

(IV) For each  $C \in [1, 2]$  there is some  $f \in \mathcal{F}_{1,2} \setminus \{0\}$  such that  $C_f = C$ ; in particular, it follows that the bounds 1 and 2 on  $C_f$  in (1.5) are the best possible ones.

Since all functions  $f$  in  $\mathcal{F}_{1,2}$  are nonnegative, the expressions on both sides of inequality (1.2) are well defined. At that, it is possible for the right-hand side, or for both sides, of (1.2) to equal  $\infty$ . In the case when the differences  $X_1, \dots, X_n$  are independent zero-mean r.v.'s, if the left-hand side of (1.2) is finite then (by Jensen's inequality)  $\mathbb{E} f(X_j) < \infty$  for each  $j \in \overline{1, n}$ , so that the right-hand side is finite as well; thus, for independent zero-mean  $X_1, \dots, X_n$ , the two sides of (1.2) are either both finite or both infinite.

A serious obstacle to overcome in order to obtain Theorem 1.1 was to understand the form of the inequality to prove, including choosing the “right” class of moment functions and, especially, developing the conjecture on what the optimal constant factor  $C_f$  in the inequality can possibly be.

## 1.2. Discussion

In this subsection, we shall

1. describe the structure of the class  $\mathcal{F}_{1,2}$  as a convex cone, which will be useful in most of the proofs, and provide examples of functions in the class  $\mathcal{F}_{1,2}$ , including the (absolute) power functions and “extreme” functions (that is, functions belonging to the extreme rays of the convex cone  $\mathcal{F}_{1,2}$ );
2. present a general approach to effective calculation of the best possible constant  $C_f$ , with further information on this constant for the power functions and “extreme” functions;
3. give an application to the concentration of measure for separately Lipschitz functions on product spaces;
4. state other corollaries of the main theorem and relate the results with the relevant ones in the literature, by von Bahr and Esseen (vBE) and other authors.

Each of these items will be presented in a separate subsubsection.

### 1.2.1. Structure of the class $\mathcal{F}_{1,2}$ and examples of functions in this class

The following proposition describes the convex-cone structure of the class  $\mathcal{F}_{1,2}$ .

**Proposition 1.2.**

- (I) A function  $f: \mathbb{R} \rightarrow \mathbb{R}$  belongs to the class  $\mathcal{F}_{1,2}$  if and only if there exists a (nonnegative, possibly infinite) Borel measure  $\gamma = \gamma_f$  on  $(0, \infty]$  such that  $\int_{(0, \infty]} (t \wedge 1) \gamma(dt) < \infty$  and

$$f(x) = \int_{(0, \infty]} \psi_t(x) \gamma(dt) \quad (1.6)$$

for all  $x \in \mathbb{R}$ , where

$$\psi_t(x) := x^2 - (|x| - t)_+^2,$$

assuming the conventions  $u_+ := 0 \vee u$ ,  $u_+^p := (u_+)^p$ ,  $u - \infty := -\infty$ , and  $(-\infty)_+ := 0$ , for all real  $u$ , so that  $\psi_\infty(x) = x^2$  for all  $x \in \mathbb{R}$ . Also,

$$\frac{1}{2t} \psi_t(x) \xrightarrow[t \downarrow 0]{} |x| \quad (1.7)$$

uniformly in  $x \in \mathbb{R}$ .

- (II) For each  $f \in \mathcal{F}_{1,2}$ , the corresponding measure  $\gamma = \gamma_f$  is unique and determined by the condition that

$$\gamma((x, \infty]) = \frac{1}{2} f''(x) \quad (1.8)$$

for all  $x \in (0, \infty)$ .

- (III) For any  $f \in \mathcal{F}_{1,2}$  and  $x \in [0, \infty)$ ,

$$f'(x) = \int_{(0, \infty]} \psi'_t(x) \gamma(dt) = 2 \int_{(0, \infty]} (x \wedge t) \gamma(dt). \quad (1.9)$$

Proposition 1.2 will be used in the proofs of most of the other results of this paper.

Note that the rays  $\mathbb{R}_+ \psi_t$  corresponding to the functions  $\psi_t$  (for  $t \in (0, \infty]$ ) are precisely the extreme rays of the convex cone  $\mathcal{F}_{1,2}$ , where  $\mathbb{R}_+ f := \{\lambda f: \lambda \in (0, \infty)\}$ , for any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ . This follows because the rays  $\mathbb{R}_+ \gamma_{\psi_t} = \mathbb{R}_+ \delta_t$  (with  $t \in (0, \infty]$ ) are precisely the extreme rays of the corresponding convex cone  $\{\gamma_f: f \in \mathcal{F}_{1,2}\}$  of measures, where  $\delta_t$  stands for the Dirac measure at the point  $t$ . (A ray  $\mathbb{R}_+ f$  of a convex cone is called extreme if, for any nonzero  $f_1$  and  $f_2$  in the cone such that  $f_1 + f_2 = f$ , both  $f_1$  and  $f_2$  must lie on the ray.)

Also, note that  $\psi_t(x) = x^2 \mathbf{I}\{|x| < t\} + (2t|x| - t^2) \mathbf{I}\{|x| \geq t\}$ , so that  $\psi_t(x)$  equals  $x^2$  for small enough  $|x|$  and is asymptotic to  $2t|x|$  as  $|x| \rightarrow \infty$ . Thus, the “extreme” function  $\psi_t$  is in a sense intermediate between the absolute powers  $|\cdot|$  and  $|\cdot|^2$ . So, by (1.6), all functions  $f \in \mathcal{F}_{1,2}$  inherit such a property. This should explain the choice of the notation  $\mathcal{F}_{1,2}$ .

Classes of moment functions similar to  $\mathcal{F}_{1,2}$  arise naturally in extremal problems in probability and statistics; see e.g. [18, 59, 45, 21, 47, 48, 49, 6, 7, 8, 51, 50, 43];  $\mathcal{F}_{1,2}$  is especially similar to the class  $\mathcal{O}_{2,3}$  considered in [21].

Let us now give some examples of functions  $f$  in  $\mathcal{F}_{1,2}$ . The “extreme” functions  $\psi_t$  have been already mentioned. Perhaps the most important members of the class  $\mathcal{F}_{1,2}$  are the power functions  $|\cdot|^p$  with  $p \in (1, 2]$ . The function  $|\cdot|$  is not in  $\mathcal{F}_{1,2}$ , since it is not in  $C^1(\mathbb{R})$ .

It is easy to construct many other kinds of examples of functions  $f \in \mathcal{F}_{1,2}$  by (i) letting  $f''$  be (on  $(0, \infty)$ ) any function, say  $g$ , which is nonnegative, nonincreasing, right-continuous, and integrable on any interval of the form  $(0, u]$ , for any  $u \in (0, \infty)$ ; then (ii) finding  $f$  on  $[0, \infty)$  as the solution to the following initial value problem:  $f(0) = f'(0) = 0$  and  $f'' = g$  on  $(0, \infty)$ ; and finally (iii) extending  $f$  to the entire real line  $\mathbb{R}$  as an even function.

E.g., taking  $g(x) = (1+x)^{p-2}$  for  $p \in (1, 2)$  and  $x \in (0, \infty)$ , one ends up with  $f(x) = \frac{1}{p(p-1)} [(1+|x|)^p - 1 - p|x|]$  for all  $x \in \mathbb{R}$ , which is asymptotic to  $\frac{1}{2}x^2$  as  $x \rightarrow 0$  and to  $\frac{1}{p(p-1)}|x|^p$  as  $|x| \rightarrow \infty$ ; if the condition  $p \in (1, 2)$  is replaced here by  $p \in (-\infty, 0) \cup (0, 1)$ , then  $f(x)$  is asymptotic to  $\frac{|x|}{1-p}$  as  $|x| \rightarrow \infty$ . Similarly one can get  $f(x) \equiv e^{-|x|} - 1 + |x|$  (by starting with  $g(x) = e^{-x}$  for  $x \in (0, \infty)$ );  $f(x) \equiv |x| - \ln(1+|x|)$  (with  $g(x) \equiv \frac{1}{(1+x)^2}$ );  $f(x) \equiv |x| \ln(1+|x|)$  (with  $g(x) \equiv \frac{1}{1+x} + \frac{1}{(1+x)^2}$ ).

Perhaps a more interesting example is the following family of functions, which are parabolic splines (and will also be used in Remark 1.5):

$$f_{\text{alt}}(x) := \frac{(|x| - x_j)^2}{2(x_j + 1)^{2/3}} + \sum_{k=0}^{j-1} \frac{[|x| - \frac{1}{2}(x_k + x_{k+1})] (x_{k+1} - x_k)}{(x_k + 1)^{2/3}} \quad (1.10)$$

if  $x_j \leq |x| < x_{j+1}$  and  $j \in \overline{0, \infty}$ , where  $x_0 := 0$ ,  $x_1$  is any positive real number, and  $x_j := q^{2^{j-1}} - 1$  for  $q := x_1 + 1$  and all  $j \in \overline{2, \infty}$ , so that  $x_{j+1} + 1 = (x_j + 1)^2$  for all  $j = 1, 2, \dots$  (we use the standard conventions  $a^{b^c} := a^{(b^c)}$  and  $\sum_{k=0}^{-1} \dots := 0$ ).

It is easy to check that  $f_{\text{alt}} \in \mathcal{F}_{1,2}$  and  $f''_{\text{alt}}(x) = (x_j + 1)^{-2/3} = (x_{j+1} + 1)^{-1/3}$  if  $x_j \leq |x| < x_{j+1}$  and  $j \in \overline{0, \infty}$ , so that the function  $f''_{\text{alt}}$  alternates between the powers  $(|\cdot| + 1)^{-2/3}$  and  $(|\cdot| + 1)^{-1/3}$ , as shown in the left panel of Figure 1. So, one might expect that the function  $f_{\text{alt}}$  alternates (far away from 0) between something like the powers  $|\cdot|^{-2/3+2} = |\cdot|^{4/3}$  and  $|\cdot|^{-1/3+2} = |\cdot|^{5/3}$ . This expectation is only partially justified.

Indeed, introduce the (instantaneous) “effective” exponent of the function  $f_{\text{alt}}$  at a point  $x \in \mathbb{R} \setminus \{0\}$  by the formula

$$p_{\text{eff}}(x) := \log_{|x|} f_{\text{alt}}(x), \quad \text{so that} \quad f_{\text{alt}}(x) = |x|^{p_{\text{eff}}(x)}.$$

The following proposition shows that the effective exponent  $p_{\text{eff}}$  eventually, “in the limit”, alternates between  $\frac{3}{2}$  (rather than the expected  $\frac{4}{3}$ ) and  $\frac{5}{3}$ . In this sense, one might say that  $f''_{\text{alt}}$  stays closer to  $(|\cdot| + 1)^{-1/3}$  than to  $(|\cdot| + 1)^{-2/3}$ , “most of the time”.

**Proposition 1.3.**

- (i)  $p_{\text{eff}}(x) = \tilde{p}_{\text{eff}}(\rho(x)) + o(1)$  as  $x \rightarrow \infty$ , where  $\tilde{p}_{\text{eff}}(r) := (2 - \frac{2}{3r}) \vee (1 + \frac{2}{3r})$  and  $\rho(x) := 2^{1-j} \log_q(x+1)$  for  $x \in (x_j, x_{j+1}]$ .
- (ii) For each  $j \in \overline{1, \infty}$ , the function  $\rho$  increases from 1 to 2 on the interval  $(x_j, x_{j+1}]$ .
- (iii) For each  $j \in \overline{1, \infty}$ , the approximate effective exponent  $\tilde{p}_{\text{eff}}(\rho(x))$  decreases from  $\frac{5}{3}$  to  $\frac{3}{2}$  and then increases back to  $\frac{5}{3}$  as  $x+1$  increases from  $x_j+1$  to  $(x_j+1)^{4/3}$  and then on to  $x_{j+1}+1 = (x_j+1)^2$ , respectively.

Part of the graph of the (exact) effective exponent  $p_{\text{eff}}$  (with  $x_1 = \frac{1}{10}$ ) is shown in the right panel of Figure 1. Recall that the  $x_j$ 's grow very fast in  $j$  for large  $j$ . Therefore, for better presentation, the horizontal axis in the right panel is nonlinearly rescaled so that the  $x_j$ 's appear equally spaced. Namely, what is actually shown here is part of the graph  $\{(\log_2 \log_q(x+1), p_{\text{eff}}(x)) : x > x_1\}$ ; note that  $\log_2 \log_q(x_j+1) = j-1$  for all  $j \in \overline{1, \infty}$ .

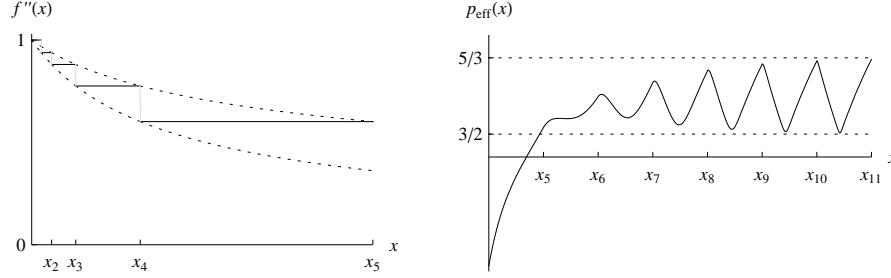


FIG 1. Left panel:  $f''$  (solid) for  $f = f_{\text{alt}}$  alternates between  $(|\cdot|+1)^{-2/3}$  (dotted) and  $(|\cdot|+1)^{-1/3}$  (dotted). Right panel: the effective exponent  $p_{\text{eff}}$  (solid) for  $f = f_{\text{alt}}$  eventually alternates between  $\frac{3}{2}$  (dotted) and  $\frac{5}{3}$  (dotted).

### 1.2.2. On the best possible constant $C_f$ in general and, in particular, for the power and extreme functions

The following proposition concerns some general properties of the constant factor  $C_f$  for nonzero  $f$  in  $\mathcal{F}_{1,2}$  except for  $f = \psi_\infty$ ; in the latter, trivial case, one has  $C_f = 1$ , as also stated in Proposition 1.6; recall that  $\psi_\infty(x) = x^2$  for all  $x \in \mathbb{R}$ .

**Proposition 1.4.** Take any  $f \in \mathcal{F}_{1,2} \setminus \{0, \psi_\infty\}$ . Let  $s_f := \inf \text{supp } \gamma$ , where  $\text{supp } \gamma$  stands for the support of the measure  $\gamma = \gamma_f$  defined in Proposition 1.2. Recall the definition of  $L_{f;s}(x)$  in (1.4). Then the following statements hold.

- (i)  $s_f \in [0, \infty)$ .
- (ii) For any  $s \in (0, s_f]$ , one has  $L_{f;s}(x) = f(s)$  for all  $x \in (0, s)$ .
- (iii) For any  $s \in (s_f, \infty)$ , one has  $L'_{f;s}(0+) > 0$  and  $L'_{f;s}(s-) < 0$ .

- (iv) For any  $s \in (0, \infty)$ , there is some (not necessarily unique)  $x_{f;s} \in (0, s)$  such that  $L_{f;s}(x)$  is nondecreasing in  $x \in (0, x_{f;s}]$  and nonincreasing in  $x \in [x_{f;s}, s)$ .
- (v) One has

$$\begin{aligned} C_f &= \sup_{s \in (s_f, \infty)} \left[ \frac{1}{f(s)} \max_{x \in (0, s)} L_{f;s}(x) \right] \\ &= \sup_{s \in (s_f, \infty)} \left[ \frac{1}{f(s)} L_{f;s}(x_{f;s}) \right] > 1. \end{aligned}$$

*Remark 1.5.* Proposition 1.4 provides for an effective maximization of  $L_{f;s}(x)$  in  $x \in (0, s)$ , for any given  $s \in (0, \infty)$ , so that  $\mathcal{L}_f(s) := \frac{1}{f(s)} \max_{x \in (0, s)} L_{f;s}(x) = \frac{1}{f(s)} L_{f;s}(x_{f;s})$  can be effectively found. In the important special case when  $f$  is a power function  $|\cdot|^p$  (with  $p \in (1, 2]$ ), one can also use the homogeneity of  $f$  in order to compute the constant  $C_f$  quite effectively, as described in Proposition 1.8. However, in general it remains to maximize  $\mathcal{L}_f(s)$  in  $s \in (s_f, \infty)$ . It appears that usually  $\mathcal{L}_f(s)$  is monotonically nondecreasing in  $s$ , if the function  $f$  is not too irregular; one “exceptional” function  $f$  for which  $\mathcal{L}_f$  lacks such a monotonicity property is a function  $f_{\text{alt}}$  of the “alternating” family described by formula (1.10). Indeed, take  $f = f_{\text{alt}}$  with  $x_1 = \frac{1}{5}$ . Then, using the Mathematica command `Maximize`, one finds that  $\mathcal{L}(\frac{107}{100}) < \mathcal{L}(\frac{106}{100})$ . One may still ask whether it is true for all  $f \in \mathcal{F}_{1,2}$  that the limit  $\mathcal{L}_f(\infty-)$  exists, and if so, whether it is true that  $\mathcal{L}_f(s) \leq \mathcal{L}_f(\infty-)$  for all  $s \in (s_f, \infty)$ , so that  $C_f$  be found as  $\mathcal{L}_f(\infty-)$ . In any case, Theorem 1.1 reduces the problem of finding the optimal constant  $C$  in (1.2) to a maximization just in two real variables,  $s$  and  $x$ , which should not usually be numerically too difficult.

Now let us provide a simple description of the constant  $C_f$  in the case when  $f$  is an “extreme” function  $\psi_t$ , representing the extreme rays of the convex cone  $\mathcal{F}_{1,2}$ :

**Proposition 1.6.** *One has  $C_{\psi_t} = 2$  for each  $t \in (0, \infty)$ , whereas  $C_{\psi_\infty} = 1$ .*

*Remark 1.7.* Proposition 1.6 might seem quite surprising: whereas, by Theorem 1.1, the range of the values of  $C_f$  over all nonzero  $f$  in the convex cone  $\mathcal{F}_{1,2}$  is the entire interval  $[1, 2]$ , the only value that  $C_f$  takes on all the extreme rays  $\mathbb{R}_+\psi_t$  (which span the cone  $\mathcal{F}_{1,2}$  in the sense of (1.6)) is 2. This suggests strong nonlinearity of the optimal constant factor  $C_f$  in  $f$ . However, as seen from the proof of Proposition 1.6, the fact that  $C_{\psi_t}$  is the same for all  $t \in (0, \infty)$  is due to a simple homogeneity property. Note also the discontinuity of  $C_{\psi_t}$  in  $t$  at  $t = \infty$ .

As mentioned earlier, for any  $p \in (1, 2]$  the power function  $|\cdot|^p$  belongs to the class  $\mathcal{F}_{1,2}$ ; for such  $p$ , consider the corresponding constant factor

$$\tilde{C}_p := C_{|\cdot|^p},$$

so that for any v-martingale  $(S_j)_{j=1}^n$

$$\mathbb{E} |S_n|^p \leq \mathbb{E} |X_1|^p + \tilde{C}_p \sum_{j=2}^n \mathbb{E} |X_j|^p. \quad (1.11)$$

Note that  $|\cdot|^2 = \psi_\infty$ , so that, by Proposition 1.6,

$$\tilde{C}_2 = 1. \quad (1.12)$$

**Proposition 1.8.**

(i) For any  $p \in (1, 2)$

$$\tilde{C}_p = \ell(p, x_p) = \max_{x \in (0,1)} \ell(p, x),$$

where

$$\ell(p, x) := L_{|\cdot|^p;1}(x) = (1-x)^p - x^p + px^{p-1} \quad (1.13)$$

for  $x \in (0, 1)$ , and  $x_p$  is the only root  $x \in (0, 1)$  of the equation

$$(1-x)^{p-1} + x^{p-1} = (p-1)x^{p-2}. \quad (1.14)$$

Moreover,  $\ell(p, x)$  is increasing in  $x \in (0, x_p)$  and decreasing in  $x \in (x_p, 1)$ , for each  $p \in (1, 2)$ .

(ii) In fact,  $x_p \in (\frac{p-1}{5}, \frac{p-1}{2}) \subset (0, \frac{1}{2})$  for all  $p \in (1, 2)$ .

(iii) Further,  $\tilde{C}_p$  is continuously (and strictly) decreasing in  $p \in (1, 2]$  from  $\tilde{C}_{1+} = 2$  to  $\tilde{C}_2 = 1$ ; furthermore,  $\tilde{C}_p$  is real-analytic in  $p \in (1, 2)$ .

(iv) The values  $\tilde{C}_p$  are algebraic for all rational  $p \in (1, 2]$ ; in particular,  $\tilde{C}_{3/2} = \sqrt{1 + \frac{1}{\sqrt{2}}} = 1.306\dots$  (with  $x_{3/2} = \frac{1}{4}(2 - \sqrt{2}) = 0.146\dots$ ).

(v) Explicit upper and lower bounds on  $\tilde{C}_p$  are given by the inequalities

$$\tilde{C}_p^{-,1} \vee \tilde{C}_p^{-,2} < \tilde{C}_p < \tilde{C}_p^{+,1} \wedge \tilde{C}_p^{+,2} \leq \tilde{C}_p^{+,2} < W_p \quad (1.15)$$

for all  $p \in (1, 2)$ , where

$$\begin{aligned} \tilde{C}_p^{-,1} &:= 2^{-p}((3-p)^p + (p-1)^{p-1}(p+1)), \\ \tilde{C}_p^{-,2} &:= 5^{-p}((6-p)^p + (p-1)^{p-1}(4p+1)), \\ \tilde{C}_p^{+,1} &:= \frac{2^{-p}}{50(3-p)}((p-1)^{p-1}(150 + 181p - 152p^2 + 21p^3) \\ &\quad + (3-p)^{p-1}(450 - 381p + 152p^2 - 21p^3)), \\ \tilde{C}_p^{+,2} &:= \frac{5^{-p}}{8(6-p)}(4(p-1)^{p-1}(12 - 35p + 94p^2 - 21p^3) \\ &\quad + (6-p)^{p-1}(288 - 15p - 94p^2 + 21p^3)), \\ W_p &:= 2^{2-p}. \end{aligned}$$

The upper bound  $W_p$  on  $\tilde{C}_p$  is exact at the endpoints of the interval  $(1, 2)$  in the sense that  $\tilde{C}_{1+} = W_{1+}$  and  $\tilde{C}_2 = \tilde{C}_{2-} = W_{2-} = W_2$ ; each of the bounds  $\tilde{C}_p^{-,1}$ ,  $\tilde{C}_p^{-,2}$ ,  $\tilde{C}_p^{+,1}$ , and  $\tilde{C}_p^{+,2}$  is also exact in the similar sense.



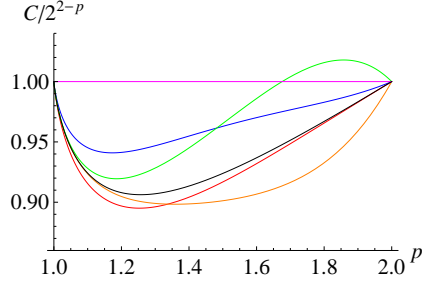


FIG 2. The ratios of  $\tilde{C}_p$  (black),  $\tilde{C}_p^{-,1}$  (red),  $\tilde{C}_p^{-,2}$  (orange),  $\tilde{C}_p^{+,1}$  (green),  $\tilde{C}_p^{+,2}$  (blue), and  $W_p$  (magenta) to  $2^{2-p}$ .

(and any r.v.'s  $X_1, \dots, X_n$ ) with  $C = \tilde{C}_1 := 1$ . From this viewpoint, there is a discontinuity of  $C_p$  at  $p = 1$ , namely,  $\tilde{C}_{1+} = 2 \neq 1 = \tilde{C}_1$ .

### 1.2.3. Application: concentration inequalities for separately Lipschitz functions on product spaces

Let  $X_1, \dots, X_n$  be independent r.v.'s with values in measurable spaces  $\mathfrak{X}_1, \dots, \mathfrak{X}_n$ , respectively. Let  $g: \mathfrak{P} \rightarrow \mathbb{R}$  be a measurable function on the product space  $\mathfrak{P} := \mathfrak{X}_1 \times \dots \times \mathfrak{X}_n$ . Let us say (cf. [9, 50]) that  $g$  is *separately Lipschitz* if it satisfies a Lipschitz type condition in each of its arguments:

$$|g(x_1, \dots, x_{i-1}, \tilde{x}_i, x_{i+1}, \dots, x_n) - g(x_1, \dots, x_n)| \leq \rho_i(\tilde{x}_i, x_i) \quad (1.16)$$

for some measurable functions  $\rho_i: \mathfrak{X}_i \times \mathfrak{X}_i \rightarrow \mathbb{R}$  and all  $i \in \overline{1, n}$ ,  $(x_1, \dots, x_n) \in \mathfrak{P}$ , and  $\tilde{x}_i \in \mathfrak{X}_i$ .

Take now any separately Lipschitz function  $g$  and let

$$Y := g(X_1, \dots, X_n).$$

Suppose that the r.v.  $Y$  has a finite mean. Then one has the following.

**Corollary 1.9.** *For each  $i \in \overline{1, n}$ , take any  $x_i \in \mathfrak{X}_i$ .*

(I) *For any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$*

$$\mathbb{E} f(Y) \leq f(\mathbb{E} Y) + \kappa_f C_f \sum_{i=1}^n \mathbb{E} f(\rho_i(X_i, x_i)), \quad (1.17)$$

where

$$\kappa_f := \sup \left\{ \frac{U_f(c, s, 0)}{U_f(c, s, a)} : s \in (0, \infty), c \in (0, \frac{s}{2}), a \in (0, c) \right\} \in [1, 2], \quad (1.18)$$

$$U_f(c, s, a) := cf(s - c + a) + (s - c)f(a - c) \quad (1.19)$$

(the above definition of  $\kappa_f$  is correct, because  $f > 0$  on  $\mathbb{R} \setminus \{0\}$  and hence  $U_f(c, s, a) > 0$  for any  $s \in (0, \infty)$ ,  $c \in (0, \frac{s}{2})$ , and  $a \in (0, c)$ ).

(II) For any  $p \in (1, 2]$

$$\mathbb{E}|Y|^p \leq |\mathbb{E}Y|^p + \tilde{\kappa}_p \tilde{C}_p \sum_{i=1}^n \mathbb{E}|\rho_i(X_i, x_i)|^p, \quad (1.20)$$

where

$$\tilde{\kappa}_p := \kappa_{|\cdot|^p} = \max_{c \in [0, 1/2]} [(c^{p-1} + (1-c)^{p-1})(c^{\frac{1}{p-1}} + (1-c)^{\frac{1}{p-1}})^{p-1}]. \quad (1.21)$$

Moreover,  $\tilde{\kappa}_p$  continuously and strictly decreases in  $p \in (1, 2]$  from 2 to 1. Furthermore, the values of  $\tilde{\kappa}_p$  are algebraic for all rational  $p \in (1, 2]$ ; in particular,  $\tilde{\kappa}_{3/2} = \frac{1}{9} \sqrt{51 + 21\sqrt{7}} = 1.14\dots$ , corresponding to  $c = \frac{1}{6}(3 - \sqrt{1 + 2\sqrt{7}}) = 0.081\dots$  in (1.21). The graph of  $\tilde{\kappa}_p$  is shown in Figure 3.

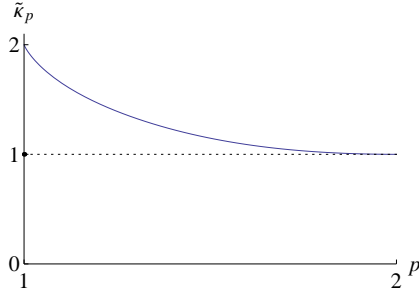


FIG 3.  $\tilde{\kappa}_p$ , solid; 1, dotted.

One can observe some similarity between  $C_f, \tilde{C}_p$  and  $\kappa_f, \tilde{\kappa}_p$ .

Thus, going from the “one-dimensional” inequality (1.2) or (1.11) for v-martingales to the “multi-dimensional” measure concentration inequality (1.17) or (1.20) entails an extra factor,  $\kappa_f$  or  $\tilde{\kappa}_p$ , whose values are between 1 and 2.

The proof of Corollary 1.9 is partly based on the following proposition, which may be of independent interest.

**Proposition 1.10.** For any zero-mean r.v.  $X$ ,  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ , and  $a \in \mathbb{R}$

$$\mathbb{E}f(X) \leq \kappa \mathbb{E}f(X + a) \quad (1.22)$$

with  $\kappa = \kappa_f$ , and  $\kappa_f$  is the best possible constant  $\kappa$  in (1.22).

In turn, the proof of Proposition 1.10 uses

**Proposition 1.11.** Take any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ ,  $s \in (0, \infty)$ , and  $c \in (0, \frac{s}{2})$ . Then  $U_f(c, s, a)$  (defined in (1.19)) is convex in  $a \in \mathbb{R}$ . Moreover,  $U_f(c, s, a)$  attains its minimum over all  $a \in \mathbb{R}$  at a unique point  $a_{f;c,s} \in [0, c)$ . In particular, for all  $t \in (0, \infty)$ ,  $s \in (0, \infty)$ , and  $c \in (0, \frac{s}{2})$

$$a_{\psi_t;c,s} = \frac{c}{s-c} (s - c - t)_+ \quad (1.23)$$

and  $\kappa_{\psi_t} = 2$ .

On the other hand, Proposition 1.11 obviously complements Corollary 1.9.

A difficulty in proving the uniqueness of the minimizer of  $U_f(c, s, a)$  in  $a$  in Proposition 1.11 is that, in general,  $U_f(c, s, a)$  is not strictly convex in  $a$ .

An example of separately Lipschitz functions  $g : \mathfrak{X}^n \rightarrow \mathbb{R}$  is given by the formula  $g(x_1, \dots, x_n) = \|x_1 + \dots + x_n\|$  for all  $x_1, \dots, x_n$  in a separable Banach space  $(\mathfrak{X}, \|\cdot\|)$ . In this case, one may take  $\rho_i(\tilde{x}_i, x_i) \equiv \|\tilde{x}_i - x_i\|$ . Thus, one obtains

**Corollary 1.12.** *Let  $X_1, \dots, X_n$  be independent random vectors in the Banach space  $(\mathfrak{X}, \|\cdot\|)$ . Let  $S_n := X_1 + \dots + X_n$ . For each  $i \in \overline{1, n}$ , take any  $x_i \in \mathfrak{X}$ . Then for any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$*

$$\mathbb{E} f(\|S_n\|) \leq f(\mathbb{E} \|S_n\|) + \kappa_f C_f \sum_{i=1}^n \mathbb{E} f(\|X_i - x_i\|). \quad (1.24)$$

Moreover, for any  $p \in (1, 2]$

$$\mathbb{E} \|S_n\|^p \leq (\mathbb{E} \|S_n\|)^p + \kappa_p \tilde{C}_p \sum_{i=1}^n \mathbb{E} \|X_i - x_i\|^p. \quad (1.25)$$

For  $p = 2$ , inequality (1.25) was obtained in [54, Theorem 4], based on an improvement of the method of Yurinskiĭ (1974) [27]; cf. [36, 37, 9], [50, Section 4], and [46, Proposition 2.5]. The proof of Corollary 1.9 is based in part on the same kind of improvement.

As can be seen from that proof, both Corollaries 1.9 and 1.12 will hold even if the separately-Lipschitz condition (1.16) is relaxed to

$$|\mathbb{E} g(x_1, \dots, x_{i-1}, \tilde{x}_i, X_{i+1}, \dots, X_n) - \mathbb{E} g(x_1, \dots, x_i, X_{i+1}, \dots, X_n)| \leq \rho_i(\tilde{x}_i, x_i). \quad (1.26)$$

Note also that in Corollaries 1.9 and 1.12 the r.v.'s  $X_i$  do not have to be zero-mean, or even to have any definable mean; at that, the arbitrarily chosen  $x_i$ 's may act as the centers, in some sense, of the distributions of the corresponding  $X_i$ 's.

Clearly, the separate-Lipschitz (sep-Lip) condition (1.16) is easier to check than a joint-Lipschitz one. Also, sep-Lip (especially in the relaxed form (1.26)) is more generally applicable. On the other hand, when a joint-Lipschitz condition is satisfied, one can generally obtain better bounds. Literature on the concentration of measure phenomenon, almost all of it for joint-Lipschitz settings, is vast; let us mention here only [31, 29, 28, 10, 30].

#### 1.2.4. Other corollaries of Theorem 1.1 and comparisons with known results

Take any  $p \in (1, 2]$ . A normed space  $(\mathfrak{X}, \|\cdot\|)$  (or, briefly,  $\mathfrak{X}$ ) is called  $p$ -uniformly smooth [1] if for some constant  $D \in (0, \infty)$  (referred to as a  $p$ -uniform smoothness constant of  $\mathfrak{X}$ ) and all  $x$  and  $y$  in  $\mathfrak{X}$  one has  $\frac{1}{2}(\|x + y\|^p + \|x - y\|^p) \leq \|x\|^p + D^p \|y\|^p$  or, equivalently,

$$\mathbb{E} \|x + Xy\|^p \leq \|x\|^p + D^p \mathbb{E} |X|^p \|y\|^p \quad (1.27)$$

for all symmetric(ally distributed) real-valued r.v.  $X$ . If  $\mathfrak{X}$  is  $p$ -uniformly smooth with a  $p$ -uniform smoothness constant  $D$ , let us say that  $\mathfrak{X}$  is  $(p, D)$ -uniformly

smooth or, simply,  $(p, D)$ -smooth. For instance, for any  $q \in [2, \infty)$  the space  $L^q(\mu)$  is  $(2, D)$ -smooth with  $D = \sqrt{q-1}$ , which is the best possible constant of the 2-uniform smoothness as long as the space  $L^q(\mu)$  is at least two-dimensional — see [46, Proposition 2.1], [1, Proposition 3], [17, Corollary 2.8].

Dual to the notion of  $(p, D)$ -uniform smoothness is that of  $(q, D^{-1})$ -uniform convexity, whose definition can be obtained by reversing the inequality sign in (1.27) and replacing there  $p$  and  $D$  by  $q$  and  $D^{-1}$ , respectively; here,  $\frac{1}{p} + \frac{1}{q} = 1$ . In particular, a result due to Ball, Carlen, and Lieb [1, Lemma 5] is that  $\mathfrak{X}$  is  $(p, D)$ -uniformly smooth iff its dual  $\mathfrak{X}^*$  is  $(q, D^{-1})$ -uniformly convex; cf. e.g. [16, 32]. Note that  $q$ -uniform convexity and  $p$ -uniform smoothness are refinements of the notions of uniform convexity and uniform smoothness, which go back to Clarkson [12] and Day [16]; cf. [25, 62]. These notions are important in functional analysis. In particular, Pisier [55] showed that every super-reflexive space is  $q$ -uniformly convex and  $p$ -uniformly smooth for some  $q$  and some  $p$ ; an earlier result due to Enflo [19] stated that  $\mathfrak{X}$  is super-reflexive iff it is isomorphic to a uniformly convex space. Among many other results, Pisier [55] also showed that the super-reflexivity is equivalent to the super-Radon-Nikodym property. Applications of the 2-uniform convexity/2-uniform smoothness to Finsler manifolds were given by Ohta [41].

It is clear that  $\mathfrak{X}$  is  $(p, D)$ -smooth iff inequality (1.2) with  $C = D^p$  and  $f = \|\cdot\|^p$  holds for all martingales (or even v-martingales)  $(S_j)_{j=1}^n$  with values in  $\mathfrak{X}$  and conditionally symmetric differences  $X_2, \dots, X_n$ ; by symmetrization, the same inequality will then hold without the conditional symmetry restriction, but with the worse constant  $C = (2D)^p$  instead of  $C = D^p$ . These considerations suggest the following.

Let us say that the space  $\mathfrak{X}$  is *completely  $(p, D)$ -smooth* if inequality (1.27) holds for all *zero-mean* real-valued r.v.'s  $X$  (and all  $x$  and  $y$  in  $\mathfrak{X}$ ). It is clear that  $\mathfrak{X}$  is completely  $(p, D)$ -smooth iff inequality (1.2) with  $C = D^p$  and  $f = \|\cdot\|^p$  holds for all martingales (or even v-martingales)  $(S_j)_{j=1}^n$  with values in  $\mathfrak{X}$ . Also, Proposition 1.8 immediately implies

**Corollary 1.13.** *Take any  $p \in (1, 2]$  and any measure  $\mu$  on any measurable space. Then the space  $L^p(\mu)$  is completely  $(p, D)$ -smooth with the best possible constant  $D = \tilde{C}_p^{1/p}$ . So, for any  $n \in \overline{2, \infty}$  and v-martingale  $(S_j)_{j=1}^n$  with values in  $L^p(\mu)$ ,*

$$\mathbb{E} \|S_n\|_p^p \leq \mathbb{E} \|X_1\|_p^p + \tilde{C}_p \sum_{j=2}^n \mathbb{E} \|X_j\|_p^p \quad (1.28)$$

(cf. (1.25)).

The above discussion suggests that the form of inequality (1.2) is rather natural in such contexts as concentration of measure, uniform smoothness, and martingales (or v-martingales). Yet, in the case when the differences  $X_1, \dots, X_n$  are independent real-valued zero-mean r.v.'s, the form of the following immediate corollary of Theorem 1.1 may be more relevant.

**Corollary 1.14.** *For any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ ,  $n \in \overline{2, \infty}$ , and (real-valued)  $v$ -martingale  $(S_j)_{j=1}^n$ ,*

$$\mathbb{E} f(S_n) \leq K \sum_{j=1}^n \mathbb{E} f(X_j) \quad (1.29)$$

with  $K = C_f$ .

However, in inequality (1.29) the constant factor  $K = C_f$  is no longer the best possible one, at least for independent zero-mean  $X_j$ 's. One way to reduce the constant is as follows. In the conditions of Corollary 1.14, rewrite the right-hand side of (1.2) with  $C = C_f$  as  $C_f \sum_{j=1}^n \mathbb{E} f(X_j) - (C_f - 1) \mathbb{E} f(X_1)$ . Then, assuming that  $\mathbb{E} f(X_1) \geq \frac{\lambda}{n} \sum_{j=1}^n \mathbb{E} f(X_j)$  for some  $\lambda \in (0, \infty)$ , one sees that the constant factor  $K = C_f$  in (1.29) can be reduced by spreading the “excess”  $C_f - 1 \geq 0$  over all the summands  $\mathbb{E} f(X_1), \dots, \mathbb{E} f(X_n)$ , to get (1.29) with

$$K = C_f - \frac{\lambda}{n} (C_f - 1) \leq C_f. \quad (1.30)$$

To develop this simple observation a bit further, let us take any  $\lambda \in (0, \infty)$  and say that a sequence  $(S_j)_{j=1}^n$  is a  $\lambda$ -good rearranged- $v$ -martingale if there are (i) some  $i \in \overline{1, n}$  such that  $\mathbb{E} f(X_i) \geq \frac{\lambda}{n} \sum_{j=1}^n \mathbb{E} f(X_j)$  and (ii) a permutation  $(j_1, \dots, j_{n-1})$  of the set  $\overline{1, n} \setminus \{i\}$  such that  $(X_i, X_{j_1}, \dots, X_{j_{n-1}})$  is the difference sequence of a  $v$ -martingale. Note that, if the differences  $X_1, \dots, X_n$  of a sequence  $(S_j)_{j=1}^n$  are independent zero-mean r.v.'s, then  $(S_j)_{j=1}^n$  is a 1-good rearranged- $v$ -martingale. (In general, a  $\lambda$ -good rearranged- $v$ -martingale does not have to be a  $v$ -martingale.) Thus, one obtains

**Corollary 1.15.** *For any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ ,  $n \in \overline{2, \infty}$ , and  $\lambda$ -good rearranged- $v$ -martingale  $(S_j)_{j=1}^n$ , inequality (1.29) holds, again with  $K$  as in (1.30).*

In the special case of the power functions  $|\cdot|^p$  (with  $p \in (1, 2)$ ) in place of general  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ , an inequality of the form (1.29) was obtained by von Bahr and Esseen (vBE) [60]:

$$\mathbb{E} |S_n|^p \leq K \sum_{j=1}^n \mathbb{E} |X_j|^p, \quad (1.31)$$

with the constant factor  $K = 2 - \frac{1}{n} = 2 - \frac{1}{n}(2 - 1)$ , which, by part (iii) of Proposition 1.8, is greater than the  $K$  in (1.30), again for  $f = |\cdot|^p$  with  $p \in (1, 2)$ . The vBE inequality (1.31) has been used in various kinds of studies, see e.g. [4, 3, 57, 39, 35, 23, 38, 5, 13, 26, 20, 2, 56], among the more recent articles.

As noted by vBE [60], the special case of inequality (1.31) (with  $K = 1$ ) when the conditional distributions of the differences  $X_i$  given  $S_{i-1}$  are symmetric for all  $i \in \overline{2, n}$  easily follows from Clarkson's inequality [12]

$$|x + y|^p + |x - y|^p \leq 2|x|^p + 2|y|^p \quad (1.32)$$

for all real  $x$  and  $y$  and all  $p \in [1, 2]$ . (As pointed out in [12], inequality (1.32) obviously implies that  $L^p$  is uniformly smooth, and in fact  $p$ -uniformly smooth.)

Actually, it is easy to see that Clarkson’s inequality (1.32) is equivalent to the symmetric case of (1.31), with  $K = 1$ .

As mentioned in [60], an inequality of the form (1.31) is not of optimal order in  $n$  for independent identically distributed real-valued zero-mean  $X_i$ ’s and may be used together with a Hölder bound such as  $\mathbb{E}|S_n|^p \leq (\mathbb{E} S_n^2)^{p/2}$ . Using similar considerations together with symmetrization and truncation, Manstavichyus [34] obtained bounds on  $\mathbb{E}|S_n|^p$  from above and below, which differ from each other by an (unspecified) factor depending only on  $p$ . The proof of Theorem 1.1 (and especially that of part (II) of Lemma 2.5) shows that near-extremal r.v.’s  $X_1, \dots, X_n$ , for which the constant  $C$  in (1.2) cannot be non-negligibly less than  $C_f$ , are as follows:  $X_1$  and  $X_2$  are independent, zero-mean, and both highly skewed in the same direction (both to the right or both to the left);  $|X_2|$  is much smaller than  $|X_1|$ ; and  $X_3, \dots, X_n$  are zero or nearly so. This suggests that the inequality (1.31) should be most useful for independent real-valued zero-mean  $X_i$ ’s when the distributions of the  $X_i$ ’s are quite different from one another and/or highly skewed and/or heavy-tailed.

Again in the case when the differences  $X_1, \dots, X_n$  are independent zero-mean r.v.’s, von Bahr and Esseen [60] made an effort to improve their constant  $K = 2 - \frac{1}{n}$  in (1.31). For such  $X_i$ ’s and the values of  $p$  in a left neighborhood of 2 such that  $D(p) := \frac{13.52}{\pi(2.6)^p} \Gamma(p) \sin \frac{\pi p}{2} = \frac{2}{\pi} \left(\frac{13}{5}\right)^{2-p} \Gamma(p) \sin \frac{\pi p}{2} < 1$ , they showed that (1.31) holds with  $K = C_p^{\text{vBE}} := \frac{1}{(1-D(p))_+}$ , assuming the convention  $\frac{1}{0} :=$

$\infty$ ; in fact, the constant factor  $C_p^{\text{vBE}}$  may improve on (i.e., may be less than) the factor  $2 - \frac{1}{n}$  only for values of  $p$  in a left neighborhood of 2 such that  $D(p) < \frac{1}{2}$ . It is stated (without proof) in [60] that  $D(p)$  decreases in  $p \in (1, 2)$  and that the mentioned left neighborhood contains the interval  $[1.6, 2]$ ; cf. Figure 4, where the von Bahr–Esseen constant factor  $2 \wedge C_p^{\text{vBE}}$  is compared with the optimal (for (1.2)) constant factor  $\tilde{C}_p$ . (There are a couple of typos in [60]: in [60, (11)], one should have  $r(2.6)^r$  instead of  $(r2.6)^r$ , and also the expression [60, (12)] for  $D(p)$  should have  $\pi(2.6)^r$  instead of  $(\pi 2.6)^r$ .)

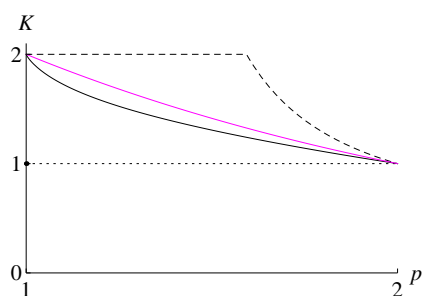


FIG 4.  $\tilde{C}_p$ , solid;  $W_p$ , magenta;  
 $2 \wedge C_p^{\text{vBE}}$ , dashed; 1, dotted.

$\|G(x)\| = \|x\|^{p-1}$ , (ii)  $G(x)x = \|x\|^p$ , and (iii)  $\|G(x) - G(y)\| \leq A\|x - y\|^{p-1}$ . The class  $\mathcal{G}_1$  was introduced by Fortet and Mourier [22]. Hoffmann–Jørgensen

The method of [60] is based on a representation of the absolute moment  $\mathbb{E}|X|^p$  of a r.v.  $X$  as a certain integral transform of the Fourier transform of the distribution of  $X$ . More general representations, for the positive-part moments  $\mathbb{E} X_+^p$ , were obtained in [11, 44].

Take now again any  $p \in (1, 2]$ . Woyczyński [61] considered the class  $\mathcal{G}_{p-1}$  of Banach spaces  $\mathfrak{X}$  defined by the following condition: there exist a map  $G: \mathfrak{X} \rightarrow \mathfrak{X}^*$  and a constant  $A = A_{p,\mathfrak{X}} \in (0, \infty)$  such that for all  $x$  and  $y$  in  $\mathfrak{X}$  one has (i)

[25] proved that  $\mathfrak{X} \in \mathcal{G}_{p-1}$  iff  $\mathfrak{X}$  is  $p$ -uniformly smooth.

Woyczyński [61] showed that inequality (1.31) holds for any independent zero-mean random vectors  $X_1, \dots, X_n$  in any Banach space  $\mathfrak{X} \in \mathcal{G}_{p-1}$ , with  $|\cdot|$  and  $K$  replaced by  $\|\cdot\|$  and  $A_{p,\mathfrak{X}}$ . As noted in [61], the space  $L^p$  is in  $\mathcal{G}_{p-1}$ , with the constant  $A = 2$ ; at that, one should take  $G(x) = x^{[p-1]} := |x|^{p-1} \text{sign } x \in L^q = (L^p)^*$  for all  $x \in L^p$ . It is not hard to see that the best possible constant  $A = A_{p,\mathfrak{X}}$  for  $\mathfrak{X} = L^p$  is

$$W_p := \sup_{u \in (-1,1)} \frac{1 - u^{[p-1]}}{(1 - u)^{p-1}} = 2^{2-p},$$

which is in agreement with the definition of  $W_p$  in part (v) of Proposition 1.8. Thus, one has (1.31) with  $K = W_p = 2^{2-p}$  for independent zero-mean differences  $X_1, \dots, X_n$ , which may be either real-valued or, equivalently, with values in  $L^p$  (in which case  $|\cdot|$  is replaced by  $\|\cdot\|_p$ ). The constant  $K = W_p$  in (1.31) is not the best possible one, even for independent zero-mean real-valued  $X_1, \dots, X_n$ , even if  $n$  is not fixed; indeed, by part (v) of Proposition 1.8,  $W_p > \tilde{C}_p$ . On the other hand, the following proposition takes place.

**Proposition 1.16.** *One has  $C_p^{\text{vBE}} > W_p$  for all  $p \in [1, 2)$ .*

So,  $C_p^{\text{vBE}} > W_p > \tilde{C}_p$  for all  $p \in (1, 2)$ . This comparison is illustrated in Figure 4.

## 2. Proofs

This section consists of four subsections. In Subsection 2.1, we shall prove 5 propositions, of the 8 ones stated in Section 1; three of these 5 propositions will be used in the proof of Theorem 1.1, in Subsection 2.4. The proof of Proposition 1.8 (which is also used in the proof of Theorem 1.1) is more involved than those of the other propositions, and it will be presented separately, in Subsection 2.2. Corollary 1.9 and the related Propositions 1.10 and 1.11 will be proved in Subsection 2.3.

The main difficulty in proving Theorem 1.1 (in distinction with discovering it) is that  $C_f$  is not an absolute constant, as it depends on  $f$ ; moreover, being the supremum of a family of ratios of linear forms in  $f$ , the factor  $C_f$  is nonlinear in  $f$  (recall here also Remark 1.7), and thus the right-hand side of the inequality (1.2) is “even more” nonlinear. Besides, the dependence of  $C_f$  on  $f$  can in general be described only implicitly. Recall also that the class  $\mathcal{F}_{1,2}$  of moment functions is an extension of the family of power functions with a small exponent  $p$  (in the interval  $(1, 2]$ ), which results in rather poor differentiability/convexity properties of the moment functions; cf. e.g. Haagerup’s case [24, 40, 42] of  $p \in (0, 2) \cup (2, 3)$  for the optimal Khinchin-type upper and lower bounds.

However, there are some favorable circumstances. First, while linearity in  $f$  is lacking, the linearity of both sides of inequality (1.2) with respect to the distribution of the v-martingale is there, and therefore it turns out to be possible to reduce inequality (1.2) to a comparison between two *quadratic* forms

in  $f$ , which in turn can be reduced to a comparison of the coefficients of these quadratic forms, as displayed in formulas (2.22), (2.24), (2.25), and (2.26). The latter comparison is already between two piecewise algebraic (even piecewise polynomial) expressions, due to the piecewise polynomial nature of the extreme functions  $\psi_t$ . The polynomials are “ $4 \times 4$ ”: each in 4 variables, of degree up to 4 in each variable. Some of these polynomials are not quite easy to analyze by hand. Moreover, the same piecewise nature of the functions  $\psi_t$  results in a very large numbers of cases to consider. After all the reductive steps described above, as well as a number of others, one still has to consider 432 such  $4 \times 4$  polynomials. While the work on each of these polynomials seems rather routine, it is a huge amount of symbolic calculations, and a good amount of numerical calculations too. In such a situation, it appears reasonable to use a computer.

A well-known result by Tarski [58, 14, 33, 15] (rooted in Sturm’s theorem) implies that systems of algebraic equations/inequalities can be solved in a completely algorithmic manner. Similar results hold for algebraic-hyperbolic polynomials (that is, polynomials in  $x, e^x, e^{-x}$ ) — as well as for certain other expressions involving inverse-trigonometric and inverse-hyperbolic functions (including the logarithmic function), whose derivatives are algebraic; this latter fact will be implicitly used in proof of Proposition 1.8. Such algorithms are implemented in Mathematica via `Reduce` and other related commands. In particular, command

```
Reduce[cond1 && cond2 && ..., {var1,var2,...}, Reals]
```

returns a simplified form of the given system (of equations and/or inequalities) `cond1, cond2, ...` over real variables `var1, var2, ...`. However, the execution of such a command may take a very long time (and/or require too much computer memory) if the given system is more than a little complicated, as is e.g. the case with the system (2.26)–(2.27) in Subsection 2.4, even after it is reduced to the 432 simpler problems, with the 432 polynomials. Hence, Mathematica will need some human help here. To keep the expressions manageable, it is important to try to simplify the expression at each step of the calculation; this can be done e.g. using Mathematica commands of the form `Assuming[cond, #//Simplify]&`. It appears that all such calculations done with a computer are, at least, as reliable and rigorous as the same calculations done by hand.

## 2.1. Proofs of Propositions 1.2, 1.3, 1.4, 1.6, and 1.16

*Proof of Proposition 1.2.* To begin, note that

$$\psi'_t(x) = 2(t \wedge x) \tag{2.1}$$

for all  $x \in [0, \infty)$  and  $t \in (0, \infty)$ . Take any  $f \in \mathcal{F}_{1,2}$ . Then, by (1.1) and the right continuity of the monotonic right derivative  $f''$  of  $f'$ , the relation (1.8)



defines a nonnegative Borel measure  $\gamma = \gamma_f$  on  $(0, \infty]$  and, by Fubini's theorem,

$$\begin{aligned} f'(x) &= \int_0^x f''(u) \, du = 2 \int_0^x du \int_{(u, \infty]} \gamma(dt) = 2 \int_{(0, \infty]} \gamma(dt) \int_0^{t \wedge x} du \\ &= 2 \int_{(0, \infty]} (t \wedge x) \gamma(dt) \end{aligned} \quad (2.2)$$

for all  $x \in [0, \infty)$ . In particular, this proves part (III) of the proposition and (taken with  $x = 1$ ) implies the condition  $\int_{(0, \infty]} (t \wedge 1) \gamma(dt) < \infty$  in part (I) of the proposition. Further, for all  $x \in [0, \infty)$  (2.2) yields

$$f(x) = \int_0^x f'(u) \, du = 2 \int_0^x du \int_{(0, \infty]} (t \wedge u) \gamma(dt) = 2 \int_{(0, \infty]} \gamma(dt) \int_0^x (t \wedge u) \, du,$$

which implies (1.6), since  $\int_0^x (t \wedge u) \, du = \frac{1}{2} \psi_t(x)$  for all  $x \in [0, \infty)$  and  $t \in (0, \infty]$ . This proves the “only if” implication in part (I) of the proposition, since the functions  $f$  and  $\psi_t$  are even.

To prove the “if” implication, assume that (1.6) holds for some nonnegative Borel measure  $\gamma$  on  $(0, \infty]$  such that  $\int_{(0, \infty]} (t \wedge 1) \gamma(dt) < \infty$  and for all  $x \in \mathbb{R}$ . In view of (2.1), the condition  $\int_{(0, \infty]} (t \wedge 1) \gamma(dt) < \infty$  implies that the integral  $\int_{(0, \infty]} \psi'_t(x) \gamma(dt)$  converges uniformly over all  $x$  in any given compact subset of the interval  $(0, \infty)$ . So, one finds that (1.6) implies (1.9), which in turn implies that  $f'$  is nondecreasing and concave on  $[0, \infty)$  (because the function  $\psi_t$  is so, for each  $t \in (0, \infty)$ ). It is also easy to see that  $f \in C^1(\mathbb{R})$ ,  $f(0) = 0$ , and  $f$  is even. Thus, it is checked that  $f \in \mathcal{F}_{1,2}$ , which completes the proof of the “if” implication in part (I) of the proposition.

It remains to prove part (II). Take indeed any  $f \in \mathcal{F}_{1,2}$ . Take also any nonnegative Borel measure  $\gamma$  on  $(0, \infty]$  such that  $\int_{(0, \infty]} (t \wedge 1) \gamma(dt) < \infty$  and (1.6) holds for all  $x \in \mathbb{R}$ . We have to show that (1.8) takes place for all  $x \in (0, \infty)$ . Take indeed any such  $x$ . Then, as has been shown, one has identities (1.9). Therefore, for any  $h \in (0, \infty)$

$$\frac{1}{2} \frac{f'(x+h) - f'(x)}{h} = \int_{(0, \infty]} r_t(x, h) \gamma(dt), \quad (2.3)$$

where  $r_t(x, h) := \frac{1}{h} [((x+h) \wedge t) - (x \wedge t)]$ , which is bounded (between 0 and 1) and converges to  $\mathbf{I}\{t > x\}$  as  $h \downarrow 0$ . So, (1.8) follows from (2.3) by dominated convergence. This completes the proof of part (II) of the proposition as well.  $\square$

*Proof of Proposition 1.3.* Part (ii) of the proposition is obvious on recalling that  $x_j = q^{2^{j-1}} - 1$  for  $j \in \overline{1, \infty}$ . Note also that  $\rho((x_j + 1)^{4/3} - 1) = \frac{4}{3}$  for  $j \in \overline{1, \infty}$ . So, to prove then part (iii), it is enough to show that  $\tilde{p}_{\text{eff}}(r)$  decreases from  $\frac{5}{3}$  to  $\frac{3}{2}$  and then increases back to  $\frac{5}{3}$  as  $r$  increases from 1 to  $\frac{4}{3}$  and then to 2, which follows because the expressions  $2 - \frac{2}{3r}$  and  $1 + \frac{2}{3r}$  are, respectively, increasing and decreasing in  $r \in [1, 2]$ , and they are equal to each other at  $r = \frac{4}{3}$ .

It remains to prove part (i) of the proposition, which is equivalent to

$$f_{\text{alt}}(x) = x^{\tilde{p}_{\text{eff}}(r)+o(1)} \quad (2.4)$$

as  $x \rightarrow \infty$ , where  $r := \rho(x) \in (1, 2]$ , so that  $x = q^{r2^{j-1}} - 1$ . In other words, it suffices to prove that the convergence (2.4) with  $x = q^{r2^{j-1}} - 1$  takes place uniformly in  $r \in (1, 2]$  as  $j \rightarrow \infty$ . Assume indeed that  $j \rightarrow \infty$  and  $x = q^{r2^{j-1}} - 1$ . Introduce  $y_j := x_j + 1$ , so that  $y_j = q^{2^{j-1}}$  for  $j = 1, 2, \dots$ . Then  $x = y_j^{r+o(1)}$ , and uniformly over all  $k \in \{0, \dots, j-1\}$  one has  $x - \frac{1}{2}(x_k + x_{k+1}) = x^{1+o(1)}$ ; moreover, if at that  $k \rightarrow \infty$  then  $x_{k+1} - x_k = x_{k+1}^{1+o(1)} = y_k^{2+o(1)}$ , which shows that the  $k$ th summand in the sum  $\sum_{k=0}^{j-1} \dots$  in (1.10) is  $(xy_k^{2-2/3})^{1+o(1)} = (y_j^r y_k^{4/3})^{1+o(1)}$  as  $k \rightarrow \infty$ . So, the sum  $\sum_{k=0}^{j-1} \dots$  in (1.10) is  $(y_j^r y_{j-1}^{4/3})^{1+o(1)} = (y_j^r y_j^{2/3})^{1+o(1)} = y_j^{r+\frac{2}{3}+o(1)}$ .

To estimate the difference  $x - x_j$ , which appears on the right-hand side of (1.10), we need to distinguish two possible cases:  $r \in [1, \frac{4}{3})$  and  $r \in [\frac{4}{3}, 2]$ . Uniformly over all  $r \in [\frac{4}{3}, 2]$  one has  $x - x_j = x^{1+o(1)} = y_j^{r+o(1)}$ , so that the term on the right-hand side of (1.10) before the sum  $\sum_{k=0}^{j-1} \dots$  is  $y_j^{2r-\frac{2}{3}+o(1)}$ , which yields  $f_{\text{alt}}(x) = y_j^{2r-\frac{2}{3}+o(1)} + y_j^{r+\frac{2}{3}+o(1)} = y_j^{(2r-\frac{2}{3}) \vee (r+\frac{2}{3})+o(1)} = y_j^{r\tilde{p}_{\text{eff}}(r)+o(1)} = x^{\tilde{p}_{\text{eff}}(r)+o(1)}$ , as in (2.4).

It remains to consider the values  $r \in [1, \frac{4}{3})$ . For such values of  $r$ , the relation  $x - x_j = x^{1+o(1)}$  no longer holds; for instance,  $x - x_j = 0$  if  $r = 1$ . However, in this case one can obviously write  $0 \leq x - x_j \leq x$  and also  $\tilde{p}_{\text{eff}}(r) = 1 + \frac{2}{3r} > 2 - \frac{2}{3r}$ . So, the term on the right-hand side of (1.10) before the sum  $\sum_{k=0}^{j-1} \dots$  is  $\leq y_j^{2r-\frac{2}{3}+o(1)} \leq y_j^{r+\frac{2}{3}+o(1)}$ , whereas still  $\sum_{k=0}^{j-1} \dots = y_j^{r+\frac{2}{3}+o(1)}$ ; so,  $y_j^{r+\frac{2}{3}+o(1)} \leq f_{\text{alt}}(x) \leq y_j^{r+\frac{2}{3}+o(1)} + y_j^{r+\frac{2}{3}+o(1)}$ , whence  $f_{\text{alt}}(x) = y_j^{r+\frac{2}{3}+o(1)} = y_j^{r\tilde{p}_{\text{eff}}(r)+o(1)} = x^{\tilde{p}_{\text{eff}}(r)+o(1)}$ , thus proving (2.4) uniformly over all  $r \in [1, \frac{4}{3})$  as well.  $\square$

*Proof of Proposition 1.4.*

(i) Since the function  $f$  is nonzero, the set  $\text{supp } \gamma$  is a nonempty subset of  $(0, \infty]$ . So,  $s_f = \inf \text{supp } \gamma \in [0, \infty]$ . If  $s_f = \infty$  then  $\text{supp } \gamma = \{\infty\}$ , which implies, in view of (1.6), that  $f = \psi_\infty$ , which contradicts the assumption on  $f$  in Proposition 1.4. This proves part (i) of the proposition.

(ii) Take any  $s \in (0, s_f]$  and  $t \in \text{supp } \gamma$ , so that  $t \in [s_f, \infty]$ . Then  $s_f > 0$  and it is straightforward to check that  $L_{\psi_t; s}(x) = \psi_t(s)$  for any  $x \in (0, s)$ . Hence, by (1.6) and (1.9),

$$L_{f; s}(x) = \int_{(0, \infty]} L_{\psi_t; s}(x) \gamma(dt) = \int_{(0, \infty]} \psi_t(s) \gamma(dt) = f(s),$$

which proves part (ii) of Proposition 1.4.

(iii) Take any  $s \in (s_f, \infty)$ . Then  $L'_{\psi_t;s}(0+) = 2(s-t)_+$  for any  $t \in (0, \infty]$ . So, by (1.9) and (1.8),

$$L'_{f;s}(0+) = \int_{(0,\infty]} L'_{\psi_t;s}(0+)\gamma(dt) = 2 \int_{(0,\infty]} (s-t)_+\gamma(dt) > 0,$$

since for any  $s \in (s_f, \infty)$  one has  $\gamma((0, s)) > 0$ . Similarly,

$$L'_{f;s}(s-) = \int_{(0,\infty]} L'_{\psi_t;s}(s-)\gamma(dt) = -2 \int_{(0,\infty]} t \mathbf{I}\{t < s\}\gamma(dt) < 0.$$

This proves part (iii) of Proposition 1.4.

(iv) In view of the rescaling identity  $L_{f;s}(x) = L_{f;s,1}(\frac{x}{s})$  with  $f_s(u) := f(su)$ , without loss of generality (w.l.o.g.)  $s = 1$ . Then part (iv) of the proposition follows by parts (ii) and (iii) and the observation that  $\ell_f(z) := L_{f;1}(1-\sqrt{z})$  is concave in  $z \in (0, 1)$ . In view of (1.6), it is enough to prove this observation for  $f = \psi_t$  with  $t \in (0, \infty]$ ; at that, by part (ii) of Proposition 1.4 and because  $s_{\psi_t} = t$ , w.l.o.g. let us assume that  $0 < t < s = 1$ . Observe that the second derivative  $\ell''_{\psi_t}(z)$  in  $z$  admits of a piecewise-algebraic expression, which may be quickly obtained by using the Mathematica command `PiecewiseExpand`. Applying then a `Reduce` command, one finds that  $\ell''_{\psi_t}(z) \leq 0$  for all  $t \in (0, 1)$  and  $z \in (0, 1)$ . Now part (iv) of Proposition 1.4 follows.

(v) Part (v) of the proposition follows by parts (i)–(iv), on recalling (1.3) and taking into account that  $L_{f;s}(0+) = f(s)$ , for all  $s \in (0, \infty)$ .

Proposition 1.4 is now completely proved.  $\square$

*Proof of Proposition 1.6.* Take any  $t \in (0, \infty]$ . That  $C_{\psi_\infty} = 1$  follows immediately by (1.3). So, w.l.o.g.  $t \in (0, \infty)$ , and then, by (1.3) and homogeneity, w.l.o.g.  $t = 1$ . Thus, it remains to show that  $C_{\psi_1} = 2$ . Take any  $s \in (s_{\psi_1}, \infty) = (1, \infty)$  and observe that  $L'_{\psi_1;s}(1) = -2(s \wedge 2) < 0$ , whereas  $L'_{\psi_1;s}(1-) = -2(s \wedge 2) + 2s \geq 0$ . Therefore, by part (iv) of Proposition 1.4,  $\max_{x \in (0,s)} L_{\psi_1;s}(x) = L_{\psi_1;s}(1) = s^2 - (s-2)_+^2$ . Now, using part (v) of Proposition 1.4, it is easy to see that  $C_{\psi_1} = \sup_{s \in (1,\infty)} \frac{s^2 - (s-2)_+^2}{s^2 - (s-1)_+^2} = \lim_{s \rightarrow \infty} \frac{s^2 - (s-2)_+^2}{s^2 - (s-1)_+^2} = 2$ .  $\square$

*Proof of Proposition 1.16.* Take any  $p \in [1, 2)$ . It suffices to show that

$$\beta(p) := (1 - D(p))2^{2-p} \stackrel{(?)}{<} 1. \quad (2.5)$$

Observe that

$$\begin{aligned} \beta'(p) &= -2^{2-p} \ln 2 + \left(\frac{26}{5}\right)^{2-p} \frac{\Gamma(p)}{\pi} \left[ 2 \left( \sin \frac{\pi p}{2} \right) \left( \ln \frac{26}{5} - (\ln \Gamma)'(p) \right) - \pi \cos \frac{\pi p}{2} \right] \\ &> -2^{2-p} \ln 2 > -2 \ln 2 > -1.4; \end{aligned}$$

the first inequality here follows because  $\cos \frac{\pi p}{2} \leq 0$ ,  $\sin \frac{\pi p}{2} > 0$ , and  $\ln \frac{26}{5} - (\ln \Gamma)'(p) \geq \ln \frac{26}{5} - (\ln \Gamma)'(2) > 0$ , taking into account that  $\ln \Gamma$  is convex and

hence  $(\ln \Gamma)'$  is increasing. It is easy to see that  $\max\{\beta(1 + \frac{i}{4}) : i \in \overline{1, 2}\} < 1 - 0.49$ . So,  $\beta(p) < \beta(1 + \frac{i}{4}) + (1.4)\frac{1}{4} < 1 - 0.49 + (1.4)\frac{1}{4} < 1$  for  $p \in [1 + \frac{i-1}{4}, 1 + \frac{i}{4}]$  and  $i \in \overline{1, 2}$ ; thus, (2.5) holds for all  $p \in [1, \frac{3}{2}]$ .

Next,

$$\beta_2(p) := 25\pi \beta''(p) 2^{p-1} = A + B(E_1 + E_2 + E_3 + E_4),$$

where

$$\begin{aligned} A &:= 50\pi \ln^2 2, \quad B := 169 \Gamma(p) \left(\frac{5}{13}\right)^p, \\ E_1 &:= 4\pi \left(\cos \frac{\pi p}{2}\right) \ln \frac{26}{5}, \quad E_2 := \kappa \sin \frac{\pi p}{2}, \\ E_3 &:= -4 \left((\ln \Gamma)'(p)^2 + (\ln \Gamma)''(p)\right) \sin \frac{\pi p}{2}, \\ E_4 &:= (\ln \Gamma)'(p) \left(-4\pi \cos \frac{\pi p}{2} + 8 \ln \frac{26}{5} \sin \frac{\pi p}{2}\right), \end{aligned}$$

and  $\kappa := \pi^2 - 4 \ln^2 2 - 4 \ln^2 \frac{13}{5} - 8 \ln 2 \ln \frac{13}{5} < 0$ , whence  $E_2 < 0$ . Also,  $E_3 < 0$ , because  $(\ln \Gamma)'' > 0$ . Let us next bound  $E_1$  and  $E_4$  from above, assuming that  $p \in [\frac{3}{2}, 2]$ . Then  $E_1 \leq 4\pi \left(\cos(\pi \frac{3}{4})\right) \ln \frac{26}{5} < -14.6$ ; also,  $(\ln \Gamma)'(p) \geq (\ln \Gamma)'(\frac{3}{2}) > 0$  and  $(\ln \Gamma)'(p) \leq (\ln \Gamma)'(2)$ , so that  $E_4 \leq (\ln \Gamma)'(2) (4\pi + 8 \ln \frac{26}{5}) < 10.9$ . Thus, for all  $p \in [\frac{3}{2}, 2]$

$$\beta_2(p) \leq 50\pi \ln^2 2 + 169 \Gamma(\frac{3}{2}) \left(\frac{5}{13}\right)^2 (-14.6 + 10.9) < -6 < 0$$

and hence  $\beta''(p) < 0$ , so that  $\beta$  is strictly concave on  $[\frac{3}{2}, 2]$ . At that,  $\beta(2) = 1$  and  $\beta'(2) = 1 - \ln 2 > 0$ ; so, (2.5) holds for all  $p \in [\frac{3}{2}, 2]$  as well.  $\square$

## 2.2. Proof of Proposition 1.8

Of the 5 parts of the proposition, the most difficult to prove are parts (iii) and (v), which are based to a certain extent on several lemmas. To state these lemmas, we need more notation. Recall the definition of  $\ell(p, x)$  in (1.13) and introduce

$$\begin{aligned} \ell_p(p, x) &:= \frac{\partial}{\partial p} \ell(p, x), \quad \ell_x(p, x) := \frac{\partial}{\partial x} \ell(p, x), \\ \ell_{x,x}(p, x) &:= \frac{\partial}{\partial x} \ell_x(p, x) = \frac{\partial^2}{\partial x^2} \ell(p, x) \end{aligned}$$

and also

$$p_x^* := \frac{1}{4}(25x + 2) \quad \text{and} \quad x_p^* := \frac{2}{25}(2p - 1),$$

so that  $x = x_p^* \iff p = p_x^*$ . Now we are ready to state the lemmas:

**Lemma 2.1.** *For all  $p \in (1, 2)$  and  $x \in (0, \frac{1}{2})$ , one has  $\ell_{x,x}(p, x) < 0$  and hence  $\ell_{x,x}(p, x) \neq 0$ .*

**Lemma 2.2.** *For all  $p \in (1, 2)$ ,*

$$B(p) := 4(p-1)^{p-1} - (6-p)^{p-1} > 0. \quad (2.6)$$

**Lemma 2.3.** *For all  $p \in (1, 2)$  and  $x \in (0, \frac{1}{2})$  such that  $x \geq x_p^*$ , one has  $\ell_x(p, x) < 0$ .*

**Lemma 2.4.** *For all  $p \in (1, 2)$  and  $x \in (0, \frac{1}{2})$  such that  $x < x_p^*$ , one has  $\ell_p(p, x) < 0$ .*

The proofs of these lemmas are deferred to the end of this subsection. Let us now consider the four parts of Proposition 1.8.

(i,ii) Take any  $p \in (1, 2)$ . Observe that  $\ell_x(p, \frac{p-1}{2}) = 2^{1-p}((p-1)^{p-1} - (3-p)^{p-1})p < 0$ , since  $p-1 < 3-p$ . On the other hand,  $\ell_x(p, \frac{p-1}{5}) = 5^{1-p}pB(p) > 0$ , by Lemma 2.2. So, any value of  $x_{f;s}$  as in part (iv) of Proposition 1.4 (for  $f = |\cdot|^p$ ) must be in the interval  $(\frac{p-1}{5}, \frac{p-1}{2}) \subset (0, \frac{1}{2})$ . By Lemma 2.1 and part (iii) of Proposition 1.4 (with  $s_f = 0$ ),  $\ell_x(p, x)$  is strictly decreasing in  $x \in (0, \frac{1}{2})$  from a positive value to a negative one. Now, in view of part (v) of Proposition 1.4, parts (i) and (ii) of Proposition 1.8 follow, taking also into account that the equation (1.14) is equivalent to  $\ell_x(p, x) = 0$ .

(iii) By part (i) of Proposition 1.8,  $x_p$  is the only root  $x \in (0, \frac{1}{2})$  of the equation  $\ell_x(p, x) = 0$ , for each  $p \in (1, 2)$ . So, by Lemma 2.1 and the implicit function theorem,  $\tilde{C}_p$  is differentiable, and even real-analytic, and hence continuous in  $p \in (1, 2)$ .

Next, by Lemma 2.3, for any  $p \in (1, 2)$  and  $x \in (0, \frac{1}{2})$  the equality  $\ell_x(p, x) = 0$  implies  $x < x_p^*$ , which in turn implies  $\ell_p(p, x) < 0$ , by Lemma 2.4. So, for any  $p \in (1, 2)$  one has  $\ell_p(p, x_p) < 0$ , whence  $\frac{d}{dp}\tilde{C}_p = \frac{d}{dp}\ell(p, x_p) = \ell_p(p, x_p) + \ell_x(p, x_p)\frac{\partial}{\partial p}x_p = \ell_p(p, x_p) < 0$ , which verifies that  $\tilde{C}_p$  is decreasing in  $p \in (1, 2)$ .

Thus, to complete the proof of part (iii) of the proposition, it remains to show that  $\tilde{C}_{1+} = 2$  and  $\tilde{C}_{2-} = 1$  (recall that  $\tilde{C}_2 = 1$ , by (1.12)). Here, consider first the case  $p \downarrow 1$ . Observe that then  $\ell(p-1, p) = (2-p)^p - (p-1)^p + p(p-1)^{p-1} \rightarrow 2$ ; on the other hand, by (1.5),  $\tilde{C}_p \leq 2$  for all  $p \in (1, 2]$ . It indeed follows that  $\tilde{C}_{1+} = 2$ . Next, for all  $x \in (0, 1)$  and  $p \in (\frac{3}{2}, 2)$ , one has  $\ell(2, x) = 1$  and  $|x^p \ln x| < |x^{p-1} \ln x| < |x^{1/2} \ln x| < \frac{2}{e} < 1$ , whence  $|\ell_p(p, x)| = |x^{p-1} + px^{p-1} \ln x - x^p \ln x + (1-x)^p \ln(1-x)| \leq |x^{p-1}| + |px^{p-1} \ln x| + |x^p \ln x| + |(1-x)^p \ln(1-x)| \leq 1 + 2 + 1 + 1 = 5$ ; so, letting  $p \uparrow 2$ , one has  $\ell(p, x) = \ell(2, x) - \int_p^2 \ell_p(r, x) dr \leq 1 + 5(2-p) \rightarrow 1$ , whence  $\limsup_{p \uparrow 2} \tilde{C}_p = \limsup_{p \uparrow 2} \ell(p, x_p) \leq 1$ . It remains to refer, again, to (1.5).

(iv) The proof of part (iv) of the proposition is straightforward.

(v) The equalities  $\tilde{C}_{1+} = W_{1+}$  and  $\tilde{C}_2 = \tilde{C}_{2-} = W_{2-} = W_2$ , and the similar equalities for the upper and lower bounds  $\tilde{C}_p^{-,1}$ ,  $\tilde{C}_p^{-,2}$ ,  $\tilde{C}_p^{+,1}$ , and  $\tilde{C}_p^{+,2}$  on  $\tilde{C}_p$  follow immediately by part (iii) of the proposition. Take now any  $p \in (1, 2)$ . Consider  $\tilde{\ell}(p, z) := \ell(p, 1 - \sqrt{z})$ , where  $z \in (0, 1)$ . By parts (i) and (ii) of Proposition 1.8,

$$\tilde{C}_p = \max_{z \in (0,1)} \tilde{\ell}(p, z) = \max_{z \in (z_1, z_2)} \tilde{\ell}(p, z),$$

where  $z_1 := z_1(p) := (\frac{3-p}{2})^2$  and  $z_2 := z_2(p) := (\frac{6-p}{5})^2$  (since the values  $\frac{p-1}{2}$  and  $\frac{p-1}{5}$  of  $x$  correspond, respectively, to the values  $z_1$  and  $z_2$  of  $z$  under the correspondence given by the formula  $x = 1 - \sqrt{z}$ .) Hence,  $\tilde{C}_p > \tilde{\ell}(p, z_1) \vee$

$\tilde{\ell}(p, z_2) = \tilde{C}_p^{-,1} \vee \tilde{C}_p^{-,2}$ , which proves the first inequality in (1.15). It follows from the proof of part (iv) of Proposition 1.4 that  $\tilde{\ell}(p, z)$  is concave in  $z \in (0, 1)$ . Also, in the proof of parts (i) and (ii) of the proposition it was observed that  $\ell_x(p, \frac{p-1}{5}) > 0 > \ell_x(p, \frac{p-1}{2})$ , which is equivalent to  $\tilde{\ell}_z(p, z_2) < 0 < \tilde{\ell}_z(p, z_1)$ , where  $\tilde{\ell}_z := \frac{\partial \tilde{\ell}}{\partial z}$ . Therefore,  $\tilde{\ell}(p, z) \leq \tilde{\ell}(p, z_1) + \tilde{\ell}_z(p, z_1)(z - z_1) < \tilde{\ell}(p, z_1) + \tilde{\ell}_z(p, z_1)(z_2 - z_1) = \tilde{C}_p^{+,1}$  and  $\tilde{\ell}(p, z) \leq \tilde{\ell}(p, z_2) + \tilde{\ell}_z(p, z_2)(z - z_2) < \tilde{\ell}(p, z_2) + \tilde{\ell}_z(p, z_2)(z_1 - z_2) = \tilde{C}_p^{+,2}$  for all  $z \in (z_1, z_2)$ , which yields the second inequality in (1.15). The third inequality in (1.15) is trivial.

So, it remains to prove the last inequality in (1.15). It is enough to show that  $\rho(p) < 0$ , where

$$\begin{aligned} \rho(p) &:= 2 \times 5^p (\tilde{C}_p^{+,2} - 2^{2-p}) \\ &= A(p) + \frac{3}{4} \frac{27-7p}{6-p} p(p-1)B(p), \\ A(p) &:= 10p(p-1)^{p-1} - 2(p-1)^p - 2^{3-p}5^p + 2(6-p)^p, \end{aligned}$$

and  $B(p)$  is as in (2.6). Observe next that  $27 - 7p \leq \frac{49}{60}(6-p)^2$ . Hence and in view of Lemma 2.2,

$$4\rho(p) \leq \tilde{\rho}(p) := 4A(p) + \frac{49}{20}(6-p)p(p-1)B(p);$$

thus, it suffices to show that  $\tilde{\rho}(p) < 0$ , which can be rewritten as  $\hat{\rho}(r) < 0$  for  $r \in (0, \frac{2}{5})$ , where

$$\hat{\rho}(r) := 16(\frac{2}{5})^{1+\frac{5}{2}r} \tilde{\rho}(1 + \frac{5}{2}r).$$

One has

$$\rho_1(s) := \hat{\rho}'(r) \frac{(1+s)^3}{r^{5r/2}} = A_1(s) + 4B_1(s)s^{\frac{5}{s+1}},$$

where

$$\begin{aligned} A_1(s) &:= 16(-62 + 2202s + 1160s^2 + 121s^3) + 80(40 + 382s + 105s^2 + 8s^3) \ln \frac{2}{1+s}, \\ B_1(s) &:= 1572 - 367s - 795s^2 - 81s^3 + (-1310s + 75s^2 + 160s^3) \ln \frac{2s}{1+s}, \end{aligned}$$

and  $s := \frac{2}{r} - 1$ , so that  $r = \frac{2}{1+s}$ , and  $r \in (0, \frac{2}{5})$  iff  $s > 4$ . Using a **Reduce** command, one finds that  $B_1(s)$  switches in sign from  $-$  to  $+$  as  $s$  increases from 4 to  $\infty$ , and the switch occurs at a certain point  $s_* = 31.4 \dots$ . With

$$\tilde{\rho}_1(s) := \frac{\rho_1(s)}{s^{5/(1+s)}B_1(s)} = \frac{A_1(s)}{s^{5/(1+s)}B_1(s)} + 4,$$

another **Reduce** command shows (in about 12 sec) that

$$\rho_2(s) := \tilde{\rho}_1'(s)B_1(s)^2 s^{(6+s)/(1+s)} \frac{(1+s)^2}{80}$$

switches in sign from  $+$  to  $-$  to  $+$  to  $-$  as  $s$  increases from 4 to  $\infty$ , and the switches occur at certain points  $s_1 = 5.2 \dots$ ,  $s_2 = 21.5 \dots$ , and  $s_3 = 42.7 \dots$ . So,  $\tilde{\rho}_1(s)$  switches from increase to decrease to increase as  $s$  increases from 4

to  $s_1 = 5.2 \dots$  to  $s_2 = 21.5 \dots$  to  $s_* = 31.4 \dots$ , and then  $\tilde{\rho}_1(s)$  switches from increase to decrease as  $s$  increases from  $s_* = 31.4 \dots$  to  $s_3 = 42.7 \dots$  to  $\infty$ . Next,  $\tilde{\rho}_1(s) < 0$  for  $s \in \{4, s_1, s_2, s_3\}$ ; also,  $\rho_1(s_*) < 0$ , whence  $\tilde{\rho}_1(s_*-) = \infty > 0$  and  $\tilde{\rho}_1(s_*+) = -\infty < 0$  (on recalling the definitions of  $\tilde{\rho}_1(s)$  and  $s_*$ ). It follows that  $\tilde{\rho}_1(s)$  switches in sign from  $-$  to  $+$  as  $s$  increases from 4 to  $s_*$ , and  $\tilde{\rho}_1 < 0$  on  $(s_*, \infty)$ . Therefore,  $\rho_1(s)$  switches in sign from  $+$  to  $-$  as  $s$  increases from 4 to  $\infty$ . Equivalently,  $\hat{\rho}'(r)$  switches in sign from  $-$  to  $+$  as  $r$  increases from 0 to  $\frac{2}{5}$ . This implies that  $\hat{\rho}(r)$  switches from decrease to increase as  $r$  increases from 0 to  $\frac{2}{5}$ . Equivalently,  $(\frac{2}{5})^p \tilde{\rho}(p)$  switches from decrease to increase as  $p$  increases from 1 to 2. Note also that  $\tilde{\rho}(1+) = \tilde{\rho}(2-) = \tilde{\rho}(2) = 0$ . So, indeed  $\tilde{\rho}(p) < 0$ , for all  $p \in (1, 2)$ . This proves part (v) and thus the entire proposition, modulo Lemmas 2.1–2.4.

*Proof of Lemma 2.1.* Introduce the new variable  $y := \frac{1-x}{x}$ , so that  $y > 1$  for  $x \in (0, \frac{1}{2})$ . Then, for any  $p \in (1, 2)$  and  $x \in (0, \frac{1}{2})$ ,

$$\begin{aligned} \ell_{x,x}(p, x) \frac{(1-x)^{2-p}}{p(p-1)} &= 1 - (2-p)y^{3-p} - (3-p)y^{2-p} \\ &< 1 - (2-p) - (3-p) = 2(p-2) < 0, \end{aligned}$$

which proves the lemma.  $\square$

*Proof of Lemma 2.2.* Take indeed any  $p \in (1, 2)$ . Note that (2.6) is equivalent to  $\tilde{B}(p) := \ln(4(p-1)^{p-1}) - \ln((6-p)^{p-1}) > 0$ . Next,  $\tilde{B}'(p) = 1 + r + \ln r$ , where  $r := \frac{p-1}{6-p}$ , so that  $\tilde{B}'(p)$  is increasing in  $p$ , and  $\tilde{B}'(2) < 0$ , which implies that  $\tilde{B}(p) < 0$  and hence  $\tilde{B}(p)$  is decreasing in  $p$ , with  $\tilde{B}(2) = 0$ . Thus, indeed  $\tilde{B}(p) > 0$ .  $\square$

*Proof of Lemma 2.3.* Throughout the proof, it is assumed that indeed  $p \in (1, 2)$  and  $x \in (0, \frac{1}{2})$ . Let

$$(D_x \ell)(p, x) := \frac{\ell_x(p, x)}{p(1-x)^{p-1}},$$

so that  $D_x \ell$  equals  $\ell_x$  in sign. Then  $\frac{\partial}{\partial x}(D_x \ell)(p, x) = (p-2)(p-1)(1-x)^{-p}x^{p-3} < 0$ , so that  $(D_x \ell)(p, x)$  decreases in  $x$ . Consider now

$$H(p) := (D_x \ell)(p, x_p^*) = (27-4p)^{1-p}(4p-2)^{p-2}(21p-23) - 1.$$

Obviously,  $H(p) < 0$  for  $p \leq \frac{23}{21}$ . Let us show that  $H(p) < 0$  for  $p \in (\frac{23}{21}, 2)$  as well. Observe that

$$\begin{aligned} H'(p) &\frac{4(27-4p)^{p-1}(2p-1)^2(4p-2)^{-p}}{21p-23} \\ &= H_1(p) := \frac{25(42p^2-92p+73)}{(27-4p)(2p-1)(21p-23)} + \ln \frac{4p-2}{27-4p}. \end{aligned}$$

Using the Mathematica command `Minimize`, one finds that  $H_1(p) > 0$  and hence  $H'(p) > 0$  for  $p \in (\frac{23}{21}, 2]$ . Since  $H(2) = 0$ , it indeed follows that  $H(p) < 0$  for

$p \in (\frac{23}{21}, 2)$  and thus for all  $p \in (1, 2)$ . So, one has  $(D_x \ell)(p, x_p^*) < 0$ . Recalling that  $(D_x \ell)(p, x)$  decreases in  $x$ , one has  $(D_x \ell)(p, x) < 0$  or, equivalently,  $\ell_x(p, x) < 0$  — provided that  $x \geq x_p^*$ .  $\square$

*Proof of Lemma 2.4.* Throughout the proof, it is assumed that indeed  $p \in (1, 2)$  and  $x \in (0, \frac{1}{2})$ . Let

$$(D_p \ell)(p, x) := \frac{\ell_p(p, x)}{-(1-x)^p \ln(1-x)} = \frac{x^{p-1}(1+(p-x)\ln x)}{-(1-x)^p \ln(1-x)} - 1,$$

$$(D_p D_p \ell)(p, x) := \frac{\partial(D_p \ell)(p, x)}{\partial p} \frac{(1-x)^p \ln(1-x)}{x^{p-1} \ln x},$$

so that  $D_p \ell$  and  $D_p D_p \ell$  equal  $\ell_p$  and  $\frac{\partial(D_p \ell)}{\partial p}$  in sign, respectively. Then  $\frac{\partial}{\partial p}(D_p D_p \ell)(p, x) = \ln(1-x) - \ln x > 0$  (since  $x \in (0, \frac{1}{2})$ ), so that  $(D_p D_p \ell)(p, x)$  increases in  $p$ . Consider now

$$(D_p D_p \ell)(p_x^*, x) = \frac{[4 + (21x + 2)\ln x] \ln(1-x) - [8 + (21x + 2)\ln x] \ln x}{4 \ln x}.$$

Observe that  $1 < p_x^* < 2 \iff \frac{2}{25} < x < \frac{6}{25}$ , and then use the Mathematica command **Reduce** to find that  $(D_p D_p \ell)(p_x^*, x) > 0$  provided that  $\frac{2}{25} < x < \frac{6}{25}$ . Similarly,  $(D_p D_p \ell)(1, x) > 0$  provided that  $0 < x \leq \frac{2}{25}$ . Thus,  $(D_p D_p \ell)(1 \vee p_x^*, x) > 0$  for all  $x \in (0, \frac{6}{25})$ . Recalling that  $(D_p D_p \ell)(p, x)$  increases in  $p$ , one has  $(D_p D_p \ell)(p, x) > 0$  for all  $p \in [1 \vee p_x^*, 2)$ . It follows that  $(D_p \ell)(p, x)$  increases in  $p \in [1 \vee p_x^*, 2)$ . Now use **Reduce** to check that  $(D_p \ell)(2, x) < 0$ , which yields  $(D_p \ell)(p, x) < 0$  or, equivalently,  $\ell_p(p, x) < 0$  for  $p \in [1 \vee p_x^*, 2)$  or, equivalently, for  $x \leq x_p^*$ .  $\square$

### 2.3. Proofs of Corollary 1.9 and Propositions 1.10 and 1.11

First in this subsection we shall prove Proposition 1.11, then Proposition 1.10, and finally Corollary 1.9.

*Proof of Proposition 1.11.* The convexity of  $U_f(c, s, a)$  in  $a \in \mathbb{R}$  follows immediately from that of  $f$ . Since  $f'$  is strictly positive and nondecreasing on  $(0, \infty)$ , it follows that  $f(\infty-) = \infty$ ; similarly (or because  $f$  is even),  $f(-\infty+) = \infty$ . So,  $U_f(c, s, a) \rightarrow \infty$  as  $|a| \rightarrow \infty$ . Therefore and by continuity, there is a minimizer of  $U_f(c, s, a)$  in  $a \in \mathbb{R}$ . Take any such minimizer, say  $a_*$ . Since  $f \in \mathcal{C}^1(\mathbb{R})$ , the partial derivative of  $U_f(c, s, a)$  in  $a$  at  $a = a_*$  is 0; that is,  $cf'(s - c + a_*) + (s - c)f'(a_* - c) = 0$ , which can be rewritten as

$$cf'(s - c + a_*) = (s - c)f'(c - a_*), \quad (2.7)$$

since  $f$  is even and hence  $f'$  is odd. Recall also that  $f'$  is strictly positive and hence nowhere zero on  $(0, \infty)$ . It follows that the arguments  $s - c + a_*$  and  $c - a_*$  of  $f'$  in (2.7) must be of the same sign; noting that the sum of these arguments is  $s > 0$ , one concludes that they must be both positive; equivalently,  $a_* \in (c - s, c)$ .



Moreover,  $f'$  is positive and nondecreasing on  $(0, \infty)$  and  $0 < c < s - c$ , so that (2.7) yields  $f'(s - c + a_*) > f'(c - a_*)$  and hence

$$s - c + a_* > c - a_*. \quad (2.8)$$

If a minimizer of  $U_f(c, s, a)$  in  $a$  is not unique, then the first two partial derivatives of  $U_f(c, s, a)$  in  $a$  are identically zero for all  $a$  in some nonempty open interval  $(a_1, a_2) \subset (c - s, c)$ . That is,  $cf'(s - c + a) = (s - c)f'(c - a)$  and  $cf''(s - c + a) + (s - c)f''(a - c) = 0$  for all  $a \in (a_1, a_2)$ . Since  $f''$  is nonnegative and even, it follows that  $f''(c - a) = f''(a - c) = 0$  for all  $a \in (a_1, a_2)$ , so that  $f'' = 0$  on the interval  $(c - a_2, c - a_1)$ . Because  $a_2 \leq c$  and  $f''$  is nonnegative and nonincreasing on  $(0, \infty)$ , one has  $f'' = 0$  on the interval  $(c - a_2, \infty)$ , so that  $f'$  is constant on the same interval. On recalling (2.8), one has  $s - c + a > c - a > c - a_2$  for any  $a \in (a_1, a_2)$ , which shows that  $f'(s - c + a) = f'(c - a)$ ; however, this contradicts the previously obtained inequality  $f'(s - c + a_*) > f'(c - a_*)$  for any minimizer  $a_*$ .

Next, the formula (1.23) for the unique minimizer of  $U_{\psi_t}(c, s, a)$  in  $a$  is easy to verify by noting that the partial derivative of  $U_{\psi_t}(c, s, a)$  in  $a$  at  $a = \frac{c}{s-c}(s - c - t)_+$  is 0. Moreover, for any real  $c$  and  $t$  such that  $c > t > 0$  one has

$$\frac{U_{\psi_1}(c, s, 0)}{U_{\psi_1}(c, s, a_{\psi_1; c, s})} \xrightarrow{s \rightarrow \infty} 2 - \frac{t}{2c}, \text{ and then } 2 - \frac{t}{2c} \xrightarrow{c \rightarrow \infty} 2, \text{ which shows that } \kappa_{\psi_t} = 2.$$

It remains to prove that the unique minimizer  $a = a_{f; c, s}$  is nonnegative. Equivalently, it remains to show that the partial derivative of  $U_f(c, s, a)$  in  $a$  is no greater than 0 at  $a = 0$ , that is,

$$cf'(s - c) \geq (s - c)f'(c). \quad (2.9)$$

By the linearity relation (1.9) and homogeneity, w.l.o.g.  $f = \psi_t$  for some  $t \in (0, \infty)$ , in which case (2.9) is equivalent to  $a_{\psi_t; c, s} \geq 0$ , and that is obvious from (1.23).  $\square$

*Proof of Proposition 1.10.* Take indeed any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ . By e.g. [52, Proposition 3.18], any zero-mean probability distribution on  $\mathbb{R} \setminus \{0\}$  is a mixture of zero-mean probability distributions on 2-point sets. Therefore, w.l.o.g. the zero-mean r.v.  $X$  takes on only two values, so that  $X = X_{c,d}$ , where  $c$  and  $d$  are positive real numbers, and  $X_{c,d}$  is a r.v. such that  $P(X_{c,d} = -c) = \frac{d}{c+d}$  and  $P(X_{c,d} = d) = \frac{c}{c+d}$ . Take now any  $c$  and  $s$  such that  $0 < c < s < \infty$ , and introduce

$$R_f(c, s, a) := \frac{U_f(c, s, 0)}{U_f(c, s, a)} = \frac{\mathbb{E} f(X_{c, s-c})}{\mathbb{E} f(X_{c, s-c} + a)}. \quad (2.10)$$

So, the best constant  $\kappa$  in (1.22) is given by a formula similar to (1.18), but with the restrictions  $c \in (0, s)$  and  $a \in \mathbb{R}$  instead of  $c \in (0, \frac{s}{2})$  and  $a \in (0, c)$ . That  $c \in (0, s)$  can be reduced to  $c \in (0, \frac{s}{2})$  follows by the symmetry relation  $R_f(c, s, a) \equiv R_f(s - c, s, -a)$  and the continuity of  $R_f(c, s, a)$  in  $c$ . Finally, the condition  $a \in \mathbb{R}$  can be reduced to  $a \in (0, c)$  by Proposition 1.11 and the continuity of  $R_f(c, s, a)$  in  $a$ .  $\square$

*Proof of Corollary 1.9.*

(I) Take indeed any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ . Consider the martingale expansion

$$Y = \mathbb{E}Y + \xi_1 + \cdots + \xi_n$$

of  $Y$  with the martingale-differences

$$\xi_i := \mathbb{E}_i Y - \mathbb{E}_{i-1} Y \quad (2.11)$$

for  $i \in \overline{1, n}$ , where  $\mathbb{E}_i$  stands for the conditional expectation given the  $\sigma$ -algebra generated by  $(X_1, \dots, X_i)$ , with  $\mathbb{E}_0 := \mathbb{E}$ . For each  $i \in \overline{1, n}$  introduce the r.v.  $\eta_i := \mathbb{E}_i(Y - Y_i)$ , where  $Y_i := g(X_1, \dots, X_{i-1}, x_i, X_{i+1}, \dots, X_n)$ ; then, in view of (1.16) or (1.26),  $|\eta_i| \leq \rho_i(X_i, x_i)$ ; because  $f(u)$  is increasing in  $|u|$ , it follows that  $f(\eta_i) \leq f(\rho_i(X_i, x_i))$  and hence  $\mathbb{E}f(\eta_i) \leq \mathbb{E}f(\rho_i(X_i, x_i))$ ; also,  $\xi_i = \eta_i - \mathbb{E}_{i-1}\eta_i$ , since the r.v.'s  $X_1, \dots, X_n$  are independent. Now (1.17) follows from Theorem 1.1 and Proposition 1.10, which latter yields  $\mathbb{E}_{i-1}f(\xi_i) \leq \kappa_f \mathbb{E}_{i-1}f(\eta_i)$  and hence  $\mathbb{E}f(\xi_i) \leq \kappa_f \mathbb{E}f(\eta_i)$ .

To check the inclusion  $\kappa_f \in [1, 2]$  in (1.18), note first that the inequality  $\kappa_f \geq 1$  follows by the continuity of  $U_f(c, s, a)$  in  $a$ , at  $a = 0$ . As for the inequality  $\kappa_f \leq 2$ , it can be rewritten as

$$U_f(c, s, 0) \leq 2U_f(c, s, a) \quad (2.12)$$

for all  $s \in (0, \infty)$ ,  $c \in (0, \frac{s}{2})$ , and  $a \in (0, c)$ , where w.l.o.g.  $f = \psi_t$  (for some  $t \in (0, \infty)$ , by (1.19) and (1.6)) and  $s = 1$  (by homogeneity). Take then indeed any  $c \in (0, \frac{1}{2})$  and  $a \in (0, c)$ . By Proposition 1.11, w.l.o.g.  $a = a_{\psi_t; c, 1}$ . Using a `Simplify` Mathematica command for  $U_{\psi_t}(c, 1, a_{\psi_t; c, 1})$  and then following with a `Reduce`, one quickly verifies that (2.12) indeed holds for  $f = \psi_t$ . This completes the proof of part (I) of Corollary 1.9.

(II) To obtain the expression in (1.21) for  $\tilde{\kappa}_p = \kappa_{|\cdot|^p}$ , note first that, by homogeneity of the power function  $f = |\cdot|^p$ , w.l.o.g.  $s = 1$ . Then solve the equation (2.7) to find the unique minimizer

$$a_* = \tilde{a}_{p; c} := a_{|\cdot|^p; c} = c - \frac{c^{1/(p-1)}}{c^{1/(p-1)} + (1-c)^{1/(p-1)}} \quad (2.13)$$

of  $\tilde{U}_p(c, a) := U_{|\cdot|^p}(c, 1, a)$  in  $a$ . Finally, substitute this minimizer for  $a$  in  $\tilde{R}_p(c, a) := \frac{\tilde{U}_p(c, 0)}{\tilde{U}_p(c, a)}$  and simplify, to show that  $\hat{r}_c(p) := \tilde{R}_p(c, \tilde{a}_{p; c})$  equals the expression under the max sign in (1.21).

The continuity of  $\tilde{\kappa}_p$  in  $p$  follows because  $\hat{r}_c(p)$  is continuous in  $p \in (1, 2]$  uniformly in  $c \in [0, \frac{1}{2}]$  (indeed, the derivative,  $\hat{r}'_c(p)$ , of  $\hat{r}_c(p)$  in  $p$  is bounded over all  $c \in [0, \frac{1}{2}]$  and all  $p$  in any compact subinterval of  $(1, 2]$ ). That  $\tilde{\kappa}_2 = 1$  is trivial. To check that  $\tilde{\kappa}_{1+} = 2$ , observe that  $\tilde{R}_p(p-1, p) \rightarrow 2$  as  $p \downarrow 1$  and recall that  $\kappa_f \leq 2$  for all  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ . The statements that the values of  $\tilde{\kappa}_p$  are algebraic for all rational  $p \in (1, 2]$  and  $\tilde{\kappa}_{3/2} = \frac{1}{9} \sqrt{51 + 21\sqrt{7}} = 1.14\dots$ , corresponding to  $c = \frac{1}{6} (3 - \sqrt{1 + 2\sqrt{7}}) = 0.081\dots$ , are straightforward to check.

It remains to prove that  $\tilde{\kappa}_p$  strictly decreases in  $p \in (1, 2]$ . To accomplish this, it is enough to show that  $\hat{r}_c(p)$  does so for each  $c \in (0, \frac{1}{2})$ , since  $\hat{r}_0(p) = \hat{r}_{1/2}(p) = 1$  for all  $p \in (1, 2]$  and  $\hat{r}_c(2) = 1$  for all  $c \in [0, \frac{1}{2}]$ . Take indeed any  $p \in (1, 2)$  and  $c \in (0, \frac{1}{2})$  and observe that  $(\ln \hat{r}_c)'(p) = r_1 + r_2 - \frac{1}{p-1}r_3$ , where

$$\begin{aligned} r_1 &:= \frac{c^{p-1} \ln c + (1-c)^{p-1} \ln(1-c)}{c^{p-1} + (1-c)^{p-1}}, \\ r_2 &:= \ln \left( c^{1/(p-1)} + (1-c)^{1/(p-1)} \right), \\ r_3 &:= \frac{c^{1/(p-1)} \ln c + (1-c)^{1/(p-1)} \ln(1-c)}{c^{1/(p-1)} + (1-c)^{1/(p-1)}}. \end{aligned}$$

Note that  $r_1 + r_2 - \frac{1}{p-1}r_3 = R_1 + R_2$ , where  $R_1 := r_1 - r_3$  and  $R_2 := r_2 + (1 - \frac{1}{p-1})r_3$ . Observe that

$$R_1 = \frac{\left( \left( \frac{1-c}{c} \right)^{p-1} - \left( \frac{1-c}{c} \right)^{\frac{1}{p-1}} \right) c^{p-1+\frac{1}{p-1}} \ln \frac{1-c}{c}}{\left( c^{\frac{1}{p-1}} + (1-c)^{\frac{1}{p-1}} \right) (c^{p-1} + (1-c)^{p-1})} < 0, \quad (2.14)$$

since  $\frac{1-c}{c} > 1$  and  $p-1 < 1 < \frac{1}{p-1}$ .

It remains to show that  $R_2 < 0$ . Consider the new variable  $b := \frac{c^{1/(p-1)}}{c^{1/(p-1)} + (1-c)^{1/(p-1)}}$ , so that  $b \in (0, \frac{1}{2})$  and  $c = \frac{b^{p-1}}{b^{p-1} + (1-b)^{p-1}}$ . Then one can check that

$$R_2 = h(b) := (p-2)(b \ln b + (1-b) \ln(1-b)) - \ln(b^{p-1} + (1-b)^{p-1}) \quad (2.15)$$

and

$$h''(b)b^{2-p}(1-b)^{2-p}(b^{p-1} + (1-b)^{p-1})^2 = h_{21}(b)h_{22}(b), \quad (2.16)$$

where

$$h_{21}(b) := \left( \frac{2-p}{b} - 1 \right) \left( \frac{b}{1-b} \right)^{2-p} + 1, \quad h_{22}(b) := \left( \frac{b}{1-b} \right)^{p-1} \left( \frac{p-1}{b} - 1 \right) - 1,$$

with  $h'_{21}(b) = (p-2)(p-1)\left(\frac{b}{1-b}\right)^{-p}(1-b)^{-3} < 0$  and  $h'_{22}(b) = (p-2) \times (p-1)\left(\frac{b}{1-b}\right)^p b^{-3} < 0$ , so that both  $h_{21}(b)$  and  $h_{22}(b)$  are decreasing in  $b$ . Since  $h_{21}(\frac{1}{2}) = 2(2-p) > 0$ , it follows that  $h_{21} > 0$  on  $(0, \frac{1}{2})$ . So,  $h''(b)$  equals  $h_{22}(b)$  in sign. Since  $h_{22}(0+) = \infty > 0$  and  $h_{22}(\frac{1}{2}) = 2(p-2) < 0$ , both  $h_{22}(b)$  and  $h''(b)$  switch from  $+$  to  $-$  as  $b$  increases from 0 to  $\frac{1}{2}$ . Therefore,  $h(b)$  switches from convexity to concavity in  $b \in (0, \frac{1}{2})$ . At that,  $h(0+) = h(\frac{1}{2}) = h'(\frac{1}{2}) = 0$ . It follows that  $h < 0$  and hence  $R_2 < 0$ . This completes the proof of part (II) and thus that of the entire Corollary 1.9.  $\square$

#### 2.4. Proof of Theorem 1.1

This proofs proceeds in reductive steps. First, the theorem is reduced to the case  $n = 2$ , which is mainly treated by Lemma 2.5. In turn, Lemma 2.5 is

largely reduced to the technical Lemma 2.6, which provides a few other rounds of reduction, one of which treated in Sublemma 2.7. We shall state these two lemmas and the sublemma just where they are needed, postponing their proofs till later in this subsection.

(I, II) By induction and conditioning, parts (I) and (II) of Theorem 1.1 follow immediately from

**Lemma 2.5.**

(I) For any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ ,  $x \in \mathbb{R}$ , and zero-mean r.v.  $Y$

$$\mathbb{E} f(x + Y) \leq f(x) + C_f \mathbb{E} f(Y).$$

(II) For any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ , one has the following: if a constant factor  $C$  is such that

$$\mathbb{E} f(X + Y) \leq \mathbb{E} f(X) + C \mathbb{E} f(Y) \quad (2.17)$$

for all independent zero-mean r.v.'s  $X$  and  $Y$ , then  $C \geq C_f$ .

We shall turn to the proof of this lemma in a moment, after the proof of parts (III) and (IV) of Theorem 1.1 is completed.

(III) Take any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ . The inequality  $C_f \geq 1$  follows by (1.3), since  $L_{f;s}(x) \rightarrow f(s)$  as  $x \downarrow 0$ . On the other hand, in view of Proposition 1.6 and (1.3), one has  $L_{\psi_t;s}(x) \leq 2\psi_t(s)$  for any  $t \in (0, \infty]$  and  $x, s$  such that  $0 < x < s < \infty$ ; so, (1.6) implies  $L_{f;s}(x) \leq 2f(s)$ , whence, by (1.3),  $C_f \leq 2$ .

(IV) Part (IV) of Theorem 1.1 follows immediately from Propositions 1.6 and 1.8.

Thus, Theorem 1.1 is proved, modulo Lemma 2.5.

*Proof of Lemma 2.5.*

(I) An argument as in the proof of Proposition 1.10 shows that w.l.o.g. the zero-mean r.v.  $Y$  takes on only two values, so that  $Y = X_{c,d}$ , where  $c$  and  $d$  are positive real numbers, and  $X_{c,d}$  is a r.v. such that  $\mathbb{P}(X_{c,d} = -c) = \frac{d}{c+d}$  and  $\mathbb{P}(X_{c,d} = d) = \frac{c}{c+d}$ . Take now any  $c$  and  $s$  such that  $0 < c < s < \infty$ , and introduce

$$g_{f;c,s}(x) := \mathbb{E} f(x + X_{c,s-c}) - f(x) \quad \text{and} \quad J_{f;c,s}(x) := \frac{g_{f;c,s}(x)}{g_{f;c,s}(0)}; \quad (2.18)$$

the latter definition is correct, because  $f > 0$  on  $\mathbb{R} \setminus \{0\}$  and hence  $g_{f;c,s}(0) = \mathbb{E} f(X_{c,s-c}) > 0$ . Observe also that  $J_{f;c,s}(-x) = J_{f;s-c,s}(x)$ . Thus, part (I) of Lemma 2.5 reduces to

$$\tilde{C}_f := \sup \{ J_{f;c,s}(x) : 0 < x < \infty, 0 < c < s < \infty \} \stackrel{(?)}{\leq} C_f. \quad (2.19)$$

Accordingly, let us take any  $c$  and  $s$  such that  $0 < c < s < \infty$ .

Using integration by parts (or, more precisely, the Fubini theorem), one has the Taylor expansion  $f(x+k) = f(x) + kf'(x) + k^2 \int_0^1 (1-z)f''(x+kz) dz$  for

all real  $x$  and  $k$ , whence

$$sg_{f;c,s}(x) = cf(x+s-c) + (s-c)f(x-c) - sf(x) \quad (2.20)$$

$$= (s-c)c \int_0^1 (1-z) [(s-c)f''(x+(s-c)z) + cf''(x-cz)] dz, \quad (2.21)$$

which is nonincreasing in  $x \in [s, \infty)$ , since  $f''$  is nonincreasing on  $(0, \infty)$ . Hence, by (2.18),  $J_{f;c,s}(x)$  is nonincreasing in  $x \in [s, \infty)$ . It follows that the condition  $0 < x < \infty$  in (2.19) can be replaced by  $0 < x < s$ . So, inequality (2.19) and thus part (I) of Lemma 2.5 follow by

**Lemma 2.6.** *For any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$  and any  $x, c, s$  such that  $0 < x < s$  and  $0 < c < s < \infty$*

$$J_{f;c,s}(x) \leq \frac{L_{f;s}(x)}{f(s)}. \quad (2.22)$$

We shall turn to the proof of this lemma in a moment, after the proof of part (II) of Lemma 2.5 is completed.

(II) To prove part (II) of Lemma 2.5, take indeed any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$ . Let  $c$  and  $s$  be as in the proof of part (I) of Lemma 2.5, so that  $0 < c < s < \infty$ . Since  $f''$  is even on  $\mathbb{R}$  and nonnegative and nonincreasing on  $(0, \infty)$ , the identity (2.21) implies that  $g_{f;c,s}(u)$  converges to a finite limit as  $u \rightarrow -\infty$ , and then so does  $J_{f;c,s}(u)$ . Let now  $a$  and  $b$  be any positive real numbers. Then

$$\frac{\mathbb{E} f(X_{a,b} + X_{c,s-c}) - \mathbb{E} f(X_{a,b})}{\mathbb{E} f(X_{c,s-c})} = \frac{b}{a+b} J_{f;c,s}(-a) + \frac{a}{a+b} J_{f;c,s}(b) \xrightarrow{a \rightarrow \infty} J_{f;c,s}(b),$$

assuming that the r.v.'s  $X_{a,b}$  and  $X_{c,s-c}$  are independent. So, the constant  $C$  in (2.17) cannot be less than  $J_{f;c,s}(b)$ , for any  $c, s, b$  such that  $0 < c < s < \infty$  and  $0 < b < \infty$ . That is,  $C$  must be no less than  $\tilde{C}_f$ , the left-hand side of (2.19).

On the other hand, by l'Hospital's rule, for any  $x \in \mathbb{R}$ ,

$$J_{f;c,s}(x) \xrightarrow{c \uparrow s} \frac{L_{f;s}(x)}{f(s)}, \quad (2.23)$$

with  $L_{f;s}(x)$  as in (1.4). So, in view of (1.3),  $\tilde{C}_f \geq C_f$ , and thus  $C \geq C_f$ . So, Lemma 2.5 is proved, modulo Lemma 2.6.  $\square$

*Proof of Lemma 2.6.* Take indeed any  $f \in \mathcal{F}_{1,2} \setminus \{0\}$  and any  $x, c, s$  such that  $0 < x < s$  and  $0 < c < s < \infty$ . Let  $\gamma := \gamma_f$  be the measure as in Proposition 1.2. Then, by (2.18), (2.20), (1.4), (1.6), and (1.9), inequality (2.22) can be rewritten as

$$\int_0^\infty \int_0^\infty \lambda_t(s; x, c) \psi_u(s) \gamma(dt) \gamma(du) \stackrel{(?)}{\leq} \int_0^\infty \int_0^\infty \mu_t(s; c) \nu_u(s; x) \gamma(dt) \gamma(du), \quad (2.24)$$

where

$$\begin{aligned}\lambda_t(s; x, c) &:= sg_{\psi_t; c, s}(x) = c\psi_t(x + s - c) + (s - c)\psi_t(x - c) - s\psi_t(x), \\ \mu_t(s; c) &:= \lambda_t(s; 0, c) = c\psi_t(s - c) + (s - c)\psi_t(c), \\ \nu_t(s; x) &:= L_{\psi_t; s}(x) = \psi_t(x - s) - \psi_t(x) + s\psi'_t(x).\end{aligned}$$

Clearly, the left-hand side of (2.24) will not change if  $\lambda_t(s; x, c)\psi_u(s)$  there is replaced by  $\lambda_u(s; x, c)\psi_t(s)$ ; a similar statement holds concerning the right-hand side of (2.24). Because of this symmetry, it is enough to prove that

$$\lambda_t(s; x, c)\psi_u(s) + \lambda_u(s; x, c)\psi_t(s) \stackrel{(?)}{\leq} \mu_t(s; c)\nu_u(s; x) + \mu_u(s; c)\nu_t(s; x) \quad (2.25)$$

for all  $u$  and  $t$  in  $(0, \infty)$  and all  $x, c, s$  such that  $0 < x < s$  and  $0 < c < s < \infty$ . Using the homogeneity relations

$$\psi_t(z) = s^2\psi_{\tilde{t}}(\tilde{z}) \quad \text{and} \quad \psi'_t(z) = s\psi'_{\tilde{t}}(\tilde{z}),$$

where  $\tilde{t} := t/s$ ,  $\tilde{z} := z/s$ ,  $t \in (0, \infty)$ , and  $z \in \mathbb{R}$ , one has  $s = 1$  w.l.o.g., so that, with

$$\Lambda_t(x, c) := \lambda_t(1; x, c), \quad \Psi_t := \psi_t(1), \quad M_t(c) := \mu_t(1; c), \quad N_t(x) := \nu_t(1; x),$$

inequality (2.25) can be further rewritten as

$$\Delta := \Delta_{u, t}(x, c) := \Lambda_t(x, c)\Psi_u + \Lambda_u(x, c)\Psi_t - M_t(c)N_u(x) - M_u(c)N_t(x) \stackrel{(?)}{\leq} 0, \quad (2.26)$$

to be proved given the restrictions

$$0 < x < 1, \quad 0 < c < 1, \quad 0 < u < \infty, \quad 0 < t < \infty. \quad (2.27)$$

Note that  $\Delta$  is piecewise-polynomial in  $x, c, u, t$ , and the restrictions (2.27) on the variables  $x, c, u, t$  are linear (or, more exactly, affine). So, as discussed in the beginning of Section 2, Mathematica commands such as **Reduce** can be used here.

We shall make several observations in order to reduce the computational complexity of the problem (2.26)–(2.27). In particular, at the end of this subsection we shall prove

**Sublemma 2.7.** *For any  $c \in (0, \frac{1}{2}]$ ,  $x \in (0, 1)$ , and  $t \in (0, \infty)$  one has  $\Lambda_t(x, c) \leq \Lambda_t(x, 1 - c)$ .*

Note also that  $M_t(c) = M_t(1 - c)$ . So, w.l.o.g.

$$\frac{1}{2} < c < 1. \quad (2.28)$$

Observe next that  $\Delta = 0$  when  $c = 1$ . Hence, it suffices to show that

$$\Delta_c := -\frac{\partial \Delta}{\partial c} \stackrel{(?)}{\leq} 0 \quad (2.29)$$

for all  $x \in (0, 1)$ ,  $c \in (\frac{1}{2}, 1)$ ,  $u \in (0, \infty)$ , and  $t \in (0, \infty)$ .

Observe also that  $\Delta_c$  is expressed in terms of the values of the functions  $\psi_u$ ,  $\psi'_u$ ,  $\psi_t$ ,  $\psi'_t$  of arguments  $z$  in the set  $Z := \{z_1, \dots, z_7\}$ , where  $(z_1, \dots, z_7) := (1, x, 1+x-c, |x-c|, 1-c, c, 1-x)$ . At that, one has one of two possible cases:

$$x \leq c \quad \text{or} \quad x > c,$$

on which the algebraic expression of  $|x-c|$  depends. In each of these two cases, the expressions of  $z_1, \dots, z_7$  in terms of  $x, c$  represent pairwise distinct affine forms in  $x, c$ . So, for any distinct  $i$  and  $j$  in the set  $\overline{1, 7}$ , the set  $\{(x, c) \in P : z_i = z_j\}$  is nowhere dense in the nonempty open set  $P = (0, 1) \times (\frac{1}{2}, 1)$ . Therefore and by the continuity of  $\psi_t$  and  $\psi'_t$ , it is enough to prove inequality (2.29) for all  $(x, c, t, u) \in (0, 1) \times (\frac{1}{2}, 1) \times (0, \infty) \times (0, \infty)$  such that  $z_i \neq z_j$  for any distinct  $i$  and  $j$  in the set  $\overline{1, 7}$ . Similarly, w.l.o.g.  $x \neq c$ . Moreover, by symmetry, w.l.o.g.

$$u < t. \tag{2.30}$$

For each  $z \in Z$ , the algebraic expressions for the values  $\psi_u(z)$ ,  $\psi'_u(z)$ ,  $\psi_t(z)$ ,  $\psi'_t(z)$  depend on the signs of  $z-u$  and  $z-t$ , respectively. Therefore, the order in which the elements of set  $Z$  go (depending on the values of  $x$  and  $c$ ) is of relevance. Overall, there are  $7! = 5040$  permutations  $\sigma$  of the set  $\{1, \dots, 7\}$ . Fortunately, comparatively few of these permutations  $\sigma$  are such that the ordering  $z_{\sigma(1)} < z_{\sigma(2)} < \dots < z_{\sigma(7)}$  may be compatible with restrictions (2.27)–(2.28); namely, there are 10 such permutations with  $x < c$  and 2 such permutations with  $x > c$  (corresponding to the orderings  $x-c < 1-x < 1-c < c < x < 1 < 1+x-c$  and  $1-x < x-c < 1-c < c < x < 1 < 1+x-c$ ), to the total of 12 permutations that may be compatible with the restrictions (2.27)–(2.28); let  $\Sigma$  denote the set of these 12 permutations. (It takes Mathematica about 11–12 sec in each of the two cases ( $x < c$  and  $x > c$ ) to select the compatible permutations.)

For each permutation  $\sigma \in \Sigma$ , the value of  $t$  may fall into one (say the  $j$ th one) of the 8 intervals  $[z_{\sigma(0)}, z_{\sigma(1)}), \dots, [z_{\sigma(7)}, z_{\sigma(8)})$ , where  $z_{\sigma(0)} := 0$  and  $z_{\sigma(8)} := \infty$ ; for each such  $j$ , the value of  $u$  (which is less than  $t$ , according to the additional restriction (2.30)) may fall either into the same  $j$ th interval or into any of the intervals to the left of it. So, for each permutation  $\sigma \in \Sigma$ , there are  $\frac{1}{2}(8 \times 9) = 36$  ways for  $t$  and  $u$  to fall into one or two of the 8 intervals. Overall, one has  $12 \times 36 = 432$  cases to consider. (In fact, we disregard the restriction  $u < t$  in any of these 432 cases when both  $u$  and  $t$  fall into the same interval  $[z_{\sigma(j)}, z_{\sigma(j+1)})$ , to make the set of all pairs  $(u, t)$  simply a rectangle in  $\mathbb{R}^2$  of the form  $[z_{\sigma(j)}, z_{\sigma(j+1)}) \times [z_{\sigma(k)}, z_{\sigma(k+1)})$ , with  $j \leq k$ .)

Using Mathematica commands **Simplify** and **Reduce** as explained above, it turns out that the mentioned purely algebraic algorithm is too slow and/or RAM-consuming. A small dose of calculus helps greatly here: we set up a preliminary test to check whether  $\Delta_c$  is convex in  $u$  and/or in  $t$ , in each case of the 432 ones; then it is enough to check that  $\Delta_c \leq 0$  when at least one of the variables  $u, t$  takes a value at an endpoint of the corresponding interval  $[z_{\sigma(j)}, z_{\sigma(j+1)})$ .

(Since  $z_{\sigma(8)}$  was defined as  $\infty$ , the right endpoint of the interval  $[z_{\sigma(j)}, z_{\sigma(j+1)})$ )

is  $\infty$  if  $j = 7$ , and one may wonder as to what the value of  $\Delta_c$  is when  $t$  equals  $\infty$  and at that  $u$  possibly equals  $\infty$  as well. This problem is resolved by observing that  $\Delta_c$  is constant in  $u \in [2, \infty)$  and in  $t \in [2, \infty)$ , which follows because, given the restrictions (2.27), one has  $Z \subset [0, 2)$ , whereas  $\psi_t(z) = z^2$  and  $\psi'_t(z) = 2z$  do not depend on  $t$  provided that  $z \in [0, 2)$  and  $t \in [2, \infty)$ .)

With these preparations, most of the 432 cases can be processed rather quickly, each taking from a fraction of a second to a few seconds to (rarely) a few minutes. Overall, it takes just about 7 minutes on a standard Core 2 Duo laptop to process all of the  $10 \times 36 = 360$  cases with  $x < c$ .

However, in each of certain 13 cases of the  $2 \times 36 = 72$  ones with  $x > c$ , it takes about 10 sec or more (or much more) for the corresponding **Reduce** command to finish; in each of these 13 more difficult cases,  $\Delta_c$  is convex neither in  $u$  nor in  $t$ . With the execution time limit set at 10 sec, Mathematica processes the  $72 - 13 = 59$  easier cases with  $x > c$  in about 3.5 minutes.

To deal with the remaining 13 cases, we set up another test, based on the following elementary observation:

- (i) if a function  $h: [A, B] \rightarrow \mathbb{R}$  is such that  $h''' \geq 0$ , then  $\max_{[A, B]} h \leq \max\{h(B), h(A), h(A) + h'(A)(B - A)\}$  — so that, if the latter max is no greater than 0, then  $h \leq 0$ ;
- (ii) if a function  $h: [A, B] \rightarrow \mathbb{R}$  is such that  $h''' \leq 0$ , then  $\max_{[A, B]} h \leq \max\{h(A), h(B), h(B) + h'(B)(A - B)\}$ .

Each of the 13 remaining cases passes test (i) (with  $h := \Delta_c$  considered as a function of  $u$ ), with the total execution time of about 4 minutes for all of the 13 cases. This completes the proof of Lemma 2.6, modulo Sublemma 2.7.  $\square$

*Proof of Sublemma 2.7.* This proof is somewhat similar to, but much simpler than, that of (2.29). Take indeed any

$$c \in (0, \tfrac{1}{2}], \quad x \in (0, 1), \quad t \in (0, \infty). \quad (2.31)$$

Let here

$$\delta := \delta_t(x, c) := \Lambda_t(x, c) - \Lambda_t(x, 1 - c).$$

As in the proof of Lemma 2.6, w.l.o.g.  $t \in (0, 2]$ . Observe that  $\delta$  is expressed in terms of the values of the function  $\psi_t$  of arguments  $z$  in the set  $Z := \{z_1, \dots, z_4\}$ , where  $(z_1, \dots, z_4) := (1 + x - c, x + c, |x - c|, |x + c - 1|)$ . At that, one has w.l.o.g. one of four possible cases:

- (i)  $x < c$  &  $x + c < 1$ ,    (ii)  $x < c$  &  $x + c > 1$ ,
- (iii)  $x > c$  &  $x + c < 1$ ,    (iv)  $x < c$  &  $x + c > 1$ ,

on which the algebraic expressions of  $|x - c|$  and  $|x + c - 1|$  depend; because of the continuity of  $\delta$ , it does not matter whether the inequalities here are strict or not. Actually, case (ii) is impossible, given the restriction  $c \in (0, \frac{1}{2}]$ . Of the  $4! = 24$  permutations  $\sigma$  of the set  $\{1, \dots, 4\}$ , there are 2 permutations that may be compatible with restrictions (2.31) in case (i), 0 such permutations in case



(ii), 3 of them in case (iii), and 1 such permutation in case (iv). Let  $\Sigma$  denote here the set of these  $2+0+3+1=6$  permutations. For each permutation  $\sigma \in \Sigma$ , the value of  $t$  may fall into one of the 5 intervals  $[z_{\sigma(0)}, z_{\sigma(1)}), \dots, [z_{\sigma(7)}, z_{\sigma(5)}]$ , where  $z_{\sigma(0)} := 0$  and  $z_{\sigma(5)} := 2$ . Overall, there are  $6 \times 5 = 30$  cases to consider. In each of these cases it is quickly checked by using the `Reduce` command that  $\delta \leq 0$ ; this takes the total of about 1 sec of computer time.  $\square$

Now Theorem 1.1 is completely proved.

## References

- [1] K. Ball, E. A. Carlen, and E. H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Invent. Math.*, 115(3):463–482, 1994.
- [2] J. Barral. Moments, continuit , et analyse multifractale des martingales de Mandelbrot. *Probab. Theory Related Fields*, 113(4):535–569, 1999.
- [3] J. Barral, J. Peyri re, and Z.-Y. Wen. Dynamics of Mandelbrot cascades. *Probab. Theory Related Fields*, 144(3-4):615–631, 2009.
- [4] D. Beliaev and S. Smirnov. Harmonic measure and SLE. *Comm. Math. Phys.*, 290(2):577–595, 2009.
- [5] G. Ben Arous, L. V. Bogachev, and S. A. Molchanov. Limit theorems for sums of random exponentials. *Probab. Theory Related Fields*, 132(4):579–612, 2005.
- [6] V. Bentkus. A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand. *Liet. Mat. Rink.*, 42(3):332–342, 2002.
- [7] V. Bentkus. An inequality for tail probabilities of martingales with differences bounded from one side. *J. Theoret. Probab.*, 16(1):161–173, 2003.
- [8] V. Bentkus. On Hoeffding’s inequalities. *Ann. Probab.*, 32(2):1650–1673, 2004.
- [9] V. Bentkus. On measure concentration for separately Lipschitz functions in product spaces. *Israel J. Math.*, 158:1–17, 2007.
- [10] S. Boucheron, O. Bousquet, G. Lugosi, and P. Massart. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560, 2005.
- [11] B. M. Brown. Characteristic functions, moments, and the central limit theorem. *Ann. Math. Statist.*, 41:658–664, 1970.
- [12] J. A. Clarkson. Uniformly convex spaces. *Trans. Amer. Math. Soc.*, 40(3):396–414, 1936.
- [13] G. Cohen and M. Lin. Extensions of the Menchoff-Rademacher theorem with applications to ergodic theory. *Israel J. Math.*, 148:41–86, 2005. Probability in mathematics.
- [14] P. J. Cohen. Decision procedures for real and  $p$ -adic fields. *Comm. Pure Appl. Math.*, 22:131–151, 1969.
- [15] G. E. Collins. Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In *Quantifier elimination and cylindrical algebraic decomposition (Linz, 1993)*, Texts Monogr. Symbol. Comput., pages 85–121. Springer, Vienna, 1998.

- [16] M. M. Day. Uniform convexity in factor and conjugate spaces. *Ann. of Math. (2)*, 45:375–385, 1944.
- [17] L. Dümbgen, S. A. van de Geer, M. C. Veraar, and J. A. Wellner. Nemirovski’s inequalities revisited. *Amer. Math. Monthly*, 117(2):138–160, 2010.
- [18] M. L. Eaton. A note on symmetric Bernoulli random variables. *Ann. Math. Statist.*, 41:1223–1226, 1970.
- [19] P. Enflo. Banach spaces which can be given an equivalent uniformly convex norm. In *Proceedings of the International Symposium on Partial Differential Equations and the Geometry of Normed Linear Spaces (Jerusalem, 1972)*, volume 13, pages 281–288 (1973), 1972.
- [20] A. H. Fan. On Markov-Mandelbrot martingales. *J. Math. Pures Appl. (9)*, 81(10):967–982, 2002.
- [21] T. Figiel, P. Hitczenko, W. B. Johnson, G. Schechtman, and J. Zinn. Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities. *Trans. Amer. Math. Soc.*, 349(3):997–1027, 1997.
- [22] R. Fortet and E. Mourier. Les fonctions aléatoires comme éléments aléatoires dans les espaces de Banach. *Studia Math.*, 15:62–79, 1955.
- [23] M. González, M. Molina, and I. M. del Puerto. Geometric growth for stochastic difference equations with application to branching populations. *Bernoulli*, 12(5):931–942, 2006.
- [24] U. Haagerup. The best constants in the Khintchine inequality. *Studia Math.*, 70(3):231–283 (1982), 1981.
- [25] J. Hoffmann-Jørgensen. On the modulus of smoothness and the  $G_\alpha$ -conditions in B-spaces (Aarhus Universitet, Matematisk Institut, Preprint Series, September 1974).
- [26] O. Johnson and R. Samworth. Central limit theorem and convergence to stable laws in Mallows distance. *Bernoulli*, 11(5):829–845, 2005.
- [27] V. V. Jurinskiĭ. Exponential estimates for large deviations. *Teor. Veroyatnost. i Primenen.*, 19:152–154, 1974.
- [28] R. Latała and K. Oleszkiewicz. Between Sobolev and Poincaré. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pages 147–168. Springer, Berlin, 2000.
- [29] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
- [30] M. Ledoux and K. Oleszkiewicz. On measure concentration of vector-valued maps. *Bull. Pol. Acad. Sci. Math.*, 55(3):261–278, 2007.
- [31] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.
- [32] J. Lindenstrauss. On the modulus of smoothness and divergent series in Banach spaces. *Michigan Math. J.*, 10:241–252, 1963.
- [33] S. Łojasiewicz. Sur les ensembles semi-analytiques. In *Actes du Congrès*

- International des Mathématiciens (Nice, 1970), Tome 2*, pages 237–241. Gauthier-Villars, Paris, 1971.
- [34] Ė. Manstavichyus. Inequalities for the  $p$ th moment,  $0 < p < 2$ , of a sum of independent random variables. *Lithuanian Math. J.*, 22(1):64–67, 1982.
  - [35] E. Manstavičius. Moments of additive functions on random permutations. *Acta Appl. Math.*, 97(1-3):119–127, 2007.
  - [36] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.
  - [37] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.
  - [38] M. Molina, M. Mota, and A. Ramos. On  $L^\alpha$ -convergence ( $1 \leq \alpha \leq 2$ ) for a bisexual branching process with population-size dependent mating. *Bernoulli*, 12(3):457–468, 2006.
  - [39] G. Morvai and B. Weiss. On universal estimates for binary renewal processes. *Ann. Appl. Probab.*, 18(5):1970–1992, 2008.
  - [40] F. L. Nazarov and A. N. Podkorytov. Ball, Haagerup, and distribution functions. In *Complex analysis, operators, and related topics*, volume 113 of *Oper. Theory Adv. Appl.*, pages 247–267. Birkhäuser, Basel, 2000.
  - [41] S.-i. Ohta. Uniform convexity and smoothness, and their applications in Finsler geometry. *Math. Ann.*, 343(3):669–699, 2009.
  - [42] I. Pinelis. From a dichotomy for images to Haagerup’s inequality (preprint), arxiv:0907.2960v2 [math.GN].
  - [43] I. Pinelis. On the Bennett-Hoeffding inequality (preprint), arxiv:0902.4058v1 [math.PR].
  - [44] I. Pinelis. Positive-part moments via the Fourier-Laplace transform. *J. Theor. Probab.*, DOI 10.1007/s10959-010-0276-9, Online First.
  - [45] I. Pinelis. Extremal probabilistic problems and Hotelling’s  $T^2$  test under a symmetry condition. *Ann. Statist.*, 22(1):357–368, 1994.
  - [46] I. Pinelis. Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.*, 22(4):1679–1706, 1994.
  - [47] I. Pinelis. Optimal tail comparison based on comparison of moments. In *High dimensional probability (Oberwolfach, 1996)*, volume 43 of *Progr. Probab.*, pages 297–314. Birkhäuser, Basel, 1998.
  - [48] I. Pinelis. Fractional sums and integrals of  $r$ -concave tails and applications to comparison probability inequalities. In *Advances in stochastic inequalities (Atlanta, GA, 1997)*, volume 234 of *Contemp. Math.*, pages 149–168. Amer. Math. Soc., Providence, RI, 1999.
  - [49] I. Pinelis. Spherically symmetric functions with a convex second derivative and applications to extremal probabilistic problems. *Math. Inequal. Appl.*, 5(1):7–26, 2002.
  - [50] I. Pinelis. On normal domination of (super)martingales. *Electron. J. Probab.*, 11:no. 39, 1049–1070, 2006.
  - [51] I. Pinelis. Exact inequalities for sums of asymmetric random variables, with applications. *Probab. Theory Related Fields*, 139(3-4):605–635, 2007.

- [52] I. Pinelis. Optimal two-value zero-mean disintegration of zero-mean random variables. *Electron. J. Probab.*, 14:no. 26, 663–727, 2009.
- [53] I. Pinelis and R. Molzon. Berry-Esséen bounds for general nonlinear statistics, with applications to Pearson’s and non-central Student’s and Hotelling’s (preprint), arXiv:0906.0177v1 [math.ST].
- [54] I. F. Pinelis and A. I. Sakhanenko. Remarks on inequalities for probabilities of large deviations. *Theory Probab. Appl.*, 30(1):143–148, 1985.
- [55] G. Pisier. Martingales with values in uniformly convex spaces. *Israel J. Math.*, 20(3-4):326–350, 1975.
- [56] H. Putter and W. R. van Zwet. Empirical Edgeworth expansions for symmetric statistics. *Ann. Statist.*, 26(4):1540–1569, 1998.
- [57] D. Surgailis. A quadratic ARCH( $\infty$ ) model with long memory and Lévy stable behavior of squares. *Adv. in Appl. Probab.*, 40(4):1198–1222, 2008.
- [58] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. RAND Corporation, Santa Monica, Calif., 1948.
- [59] S. A. Utev. Extremal problems in moment inequalities. In *Limit theorems of probability theory*, volume 5 of *Trudy Inst. Mat.*, pages 56–75, 175. “Nauka” Sibirsk. Otdel., Novosibirsk, 1985.
- [60] B. von Bahr and C.-G. Esseen. Inequalities for the  $r$ th absolute moment of a sum of random variables,  $1 \leq r \leq 2$ . *Ann. Math. Statist.*, 36:299–303, 1965.
- [61] W. A. Woyczyński. Random series and laws of large numbers in some Banach spaces. *Theory Probab. Appl.*, 18:350–355, 1973.
- [62] W. A. Woyczyński. Geometry and martingales in Banach spaces. In *Probability—Winter School (Proc. Fourth Winter School, Karpacz, 1975)*, pages 229–275. Lecture Notes in Math., Vol. 472. Springer, Berlin, 1975.